ASPECTS OF MITOCHONDRIAL DNA MUTATIONS IN RELATION TO HUMAN MALE FERTILISING POTENTIAL

A thesis
submitted in fulfilment
of the requirements for the degree
of
Doctor of Philosophy
in the
Department of Zoology

by

Andrew John Holyoake

University of Canterbury
1998

TABLE OF CONTENTS

ABSTRACT

Fertilisation is a chance event dependent on, amongst other things, the number, motility, and physical characteristics of sperm. There is a strong link between ATP production and sperm motility. The mitochondrially encoded OXPHOS ATP production pathway provides ATP for sperm motility. Thirteen polypeptides involved in OXPHOS are encoded in mtDNA. Therefore, the aim of this study was to assess the influence of point mutations and small deletions in mtDNA on human male fertilising potential. Two hundred and fifty two semen samples were screened for mutations in the ATPase 6 and 8, and COII genes using polymerase chain reaction (PCR), followed by single-strand conformation polymorphism (SSCP) analysis of restriction digests of two overlapping amplicons. ATPase 6 and 8 gene analysis of 210 semen samples identified three singly occurring known homoplasmic polymorphisms (G → A 8860, G → A 8856, and G → A 8839); one polymorphism found in 13 samples that created a known *Hae*II RFLP; and a novel heteroplasmic point mutation (T → C 8821) found in the immature sperm from two semen samples from an oligo-astheno-teratozoospermic individual. Based on amino acid homologies and protein structure this new heteroplasmic mutation is predicted to be potentially pathogenic. COII gene analysis of 223 semen samples identified two previously described, singly occurring homoplasmic polymorphisms (C → T 7476, and G → A 7853); three novel singly occurring homoplasmic polymorphisms (T → C 8077, G → A 7789, and G → A 7754); two known homoplasmic polymorphisms found in multiple samples (9 bp deletion, 7 samples; G → A 8251, 11 samples) and one novel heteroplasmic homopolymeric tract insertion in the COII/tRNA^{Lys} intergenic spacer. Of the multiply occurring variants found in these genes only a 9 bp deletion from the COII/tRNA^{Lys} intergenic spacer of 7 semen samples could be associated with asthenozoospermia. This association was negated when disease and cultural factors were considered.

In addition to the characterisation of true heteroplasmy, one case of pseudogene-created heteroplasmy was characterised in the COII amplicon from one semen sample. During the analysis of this known pseudogene-created heteroplasmy, two other novel pseudogenes paralogous to the mitochondrial ND4/ND5 genes and the HV1 region of

the D-loop were isolated. All three pseudogenes were characterised and the times of pseudogene insertion were calculated. The COII and ND4/ND5 pseudogenes are predicted to have inserted into nuclear DNA since the time of the last common ancestor of humans and chimpanzees, whilst the HV1 pseudogene is predicted to have an ancient origin. These pseudogenes may act as useful markers for molecular evolution.

Chapter 1

GENERAL INTRODUCTION

Mitochondrial DNA

The human mitochondrial DNA (mtDNA) genome is a covalently closed circular molecule normally consisting of 16,569 nucleotide pairs. It encodes 2 rRNAs, 22 tRNAs, and 13 of the 67 subunits of the mitochondrial respiratory chain and oxidative phosphorylation (OXPHOS) system: 7 subunits of NADH CoQ reductase (complex I), Cytochrome b (complex III), subunits I, II, and III of Cytochrome *c* Oxidase (complex IV), and subunits 6 and 8 of the F_0F_1 ATP synthase (Fig. 1.1) (Anderson *et al.*, 1981). First sequenced in its entirety in 1981 from a mixture of human placenta (European origin) and HeLa cells, the initial human mtDNA sequence has since been deemed a worldwide reference, called the Cambridge sequence (after the locality where the sequencing was performed). This sequence has since been revised (Marzuki *et al.*, 1991). Within this thesis, reference to the Cambridge sequence, inplies the revised sequence.

Mitochondrial DNA is predominantly maternally inherited. The reasons for this are still debated but could include dilution effects (oocytes have up to 200 fold more copies of mtDNA than sperm), destruction of the mitochondrion-containing sperm mid piece (Cummins, 1997) or preferential inactivation of sperm mtDNA in the oocyte cytoplasm (Kaneda *et al.*, 1995). However, paternal transmission has been demonstrated in a number of species (e.g. Gyllensten *et al.*, 1991, Magoulas and Zouros, 1993) with mussels (*Mytilus* sp.) having almost uniform biparental inheritance (Zouros *et al.*, 1992). In sea urchins, the sperm mitochondria actively participate in embryogenesis up to the 8-cell stage (Anderson and Perotti, 1975). In mammals this has never been demonstrated although there is a requirement in the developing embryo for extra-nuclear paternal input (e.g. Sathananthan *et al.*, 1996). The lack of mixing of the paternal and maternal mtDNA at fertilisation (predominant uniparental inheritance) means that mtDNAs do not recombine, although the machinery for recombination is present (Lightowlers *et al.*, 1997). Because of the lack of recombination almost all mtDNA variation is the product of random, sequential mutations along maternal lineages.

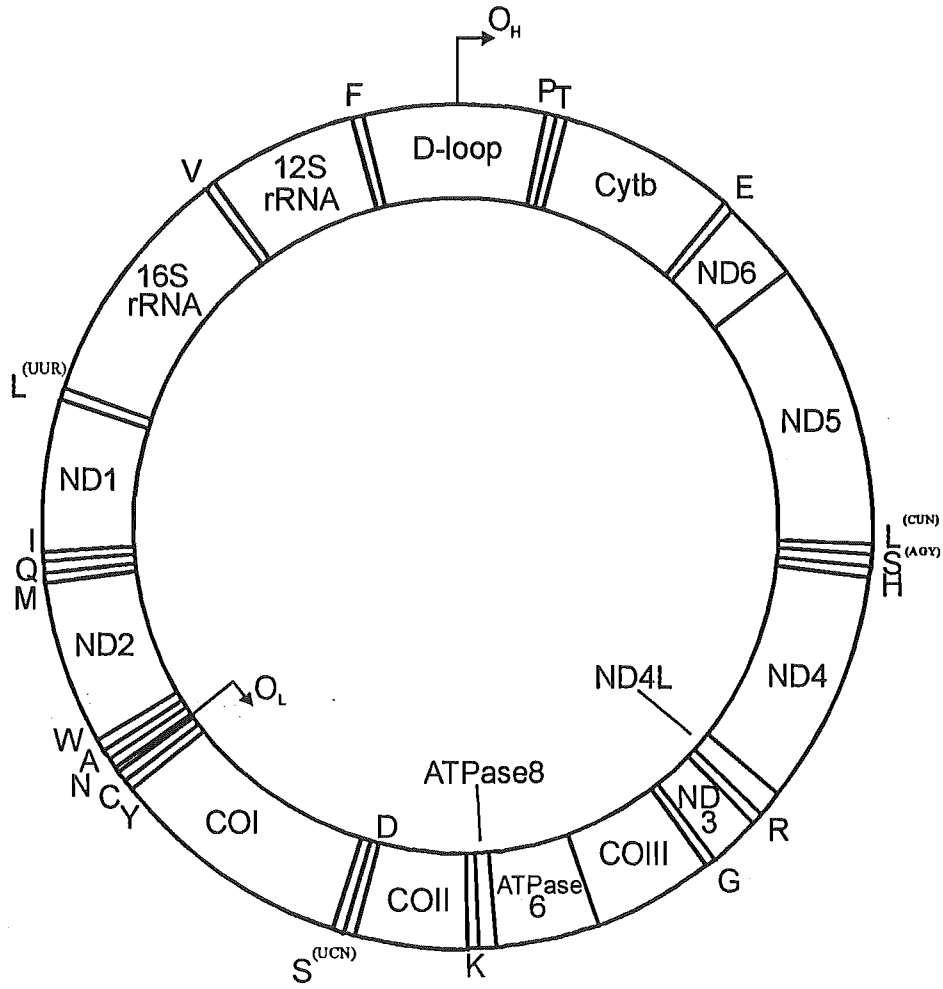


Figure 1.1. Diagrammatical representation of the human mtDNA molecule. The protein encoding genes are 7 subunits of NADH CoQ reductase (ND1, 2, 3, 4, 4L, 5, 6), Cytochrome b (Cyt b), subunits I, II, and III of Cytochrome *c* Oxidase (COI, II, III), and subunits 6 and 8 of the F_0F_1 ATP synthase (ATPase 6, 8). Transfer RNAs are refered to by their single letter codes. The iso-acceptors of serine and leucine are specified. Ribosomal RNAs are refered to by their size in Svedberg units. Origins of replication for the heavy (O_H) and light (O_L) strands are shown. No attempt has been made to specify the coding strand of each gene. Where relevant this is mentioned in the text (adapted from Wallace, 1993).

Mitochondrial DNA is particularly prone to mutations[1] since the mtDNA is not packed with histone-like proteins (Torroni and Wallace, 1994). Mitochondria have a low efficiency DNA repair system and mtDNA are attached to the mitochondrial inner membrane in close proximity to OXPHOS, where many mutagenic oxygen free radicals are generated (Bandy and Davison, 1990). As a result of the proximity of the mtDNA to OXPHOS, the

[1] The terminology defining 'changes' can be ambiguous. Throughout this thesis events that cause variation are termed mutational events, hence the variations they cause are mutations. This is not intended to imply all mutations have deleterious effects. Where mutations have been shown previously to be population variants, with no known deleterious effects they are refered to as polymorphisms.

mutation rate[2] of mtDNA, whilst not uniform, is generally 10 to 17 times higher than nuclear DNA (including pseudogenes) (Shoffner *et al.*, 1990). Paradoxically, mtDNA is not designed to accommodate a higher mutation rate. Whilst an altered genetic code including frequent use of a 'four-way pairing system' for codon assignment (Avise, 1991) may be an advantage in a high mutation system, the lack of intergenic spacers and introns to act as mutation 'sinks' is not. The mtDNA molecule has been streamlined and compacted through evolution by the loss of intergenic spacers and introns and the transfer of functional genes to the nuclear genome (Wallace, 1986). This may have provided selective advantage, as small sized mtDNA molecules will replicate faster and hence enhance fitness both at the intracellular and organismal level (Avise, 1991).

One intergenic non-coding region not lost over time is the displacement loop (D-loop) (Fig. 1.1). The D-loop contains three relatively invariant conserved sequence blocks (CSB) and two hypervariable regions (HV1 and 2) and is thought to help regulate transcription and replication (Saccone *et al.*, 1991). The lack of coding and general functional constraints in HV1 and 2 have allowed them to evolve at a very high rate to the extent that they have been investigated as an alternative to DNA fingerprinting for between-individual identification (Piercy *et al.*, 1993). This high rate of mutation and the lack of recombination has allowed mtDNA to be used as a marker for population variation and species evolution (for recent review see Cavalli-Sforza, 1998).

Mitochondrial biogenesis

Mitochondrial DNA is a double stranded molecule. The two strands differ markedly in their base composition. The heavy, (H)-strand contains most of the guanine (G) residues whilst the light (L)-strand contains most of the cytosine (C) residues. The H-strand functions as the template for the two rRNAs, 12 of the polypeptides, and 14 of the tRNAs whilst the L-strand is the template for the ND6 polypeptide and 8 tRNAs (Ojala *et al.*, 1981). MtDNA uses two replication origins; O_L and O_H. Replication from O_H is initiated first by prior RNA transcription from the light strand propmotor (LSP) using an mtRNA polymerase in association with a mitochondrial transcription factor (mtTFA). This RNA molecule binds strongly to two conserved sequence blocks (CSBI,

[2] Whilst it is acknowledged that the mutation rate and the mutation fixation rates are frequently different, the term mutation rate is used generally to cover both except where specified.

II) prior to O_H. The non-bound RNA is clipped by an MRP Rnase. The bound RNA then acts as a primer for H-strand replication. With the progression of H-strand synthesis (clockwise on Figure 1.1), the parental H-strand is displaced. L-strand synthesis is initiated in an anti-clockwise direction, when O_L is exposed (Wong and Clayton, 1986). L-strand synthesis proceeds back along the free H-strand, thus replication is bidirectional but asynchronous (for review see Shadel and Clayton, 1997)

Transcription is also bidirectional with promoters for both the H and L-strands in the D-loop facing in opposite directions. The H-strand promoter (P_H) reads anticlockwise whilst the L-strand promoter (P_L) reads clockwise (Montoya *et al.*, 1981). Both transcripts encompass the entire genome. As tRNAs are formed they fold and are cleaved out (Ojala *et al.*, 1981) releasing mRNAs and rRNAs in the process. Messenger RNA poly-A tails are added posttranscriptionally (Anderson *et al.*, 1981). As mtDNA uses a different genetic code to nuclear DNA (Anderson *et al.*, 1981) mitochondrial mRNA cannot be translated in the cytoplasm.

Mitochondrial DNA as a marker for human population variation

As a consequence of random mutations along defined lineages, human populations now harbour a wide variety of ethnic-specific changes. This has given rise to defined haplotypes which associate with specific populations or ethnic groups. However, mutations are random, continuous and reversible meaning that not all individuals with one change will belong to the same haplotype. Population variants fall into two categories. (1) Mutations can be neutral, silent changes that have no bearing on amino acid sequence or the functional properties of mtDNA. These changes in populations are termed polymorphisms. (2) Deleterious mutations, however, can also become fixed in populations if they are expressed later in life and have no affect on the reproductive fitness of an individual. As more and more sequences are compared between and within populations, more discriminatory markers are found and sub-haplotypes are created within existing haplotypes. Haplotype creation, essentially the classification of discriminatory markers, has been made possible with the help of computer aided phylogenetic analysis based on the pairwise comparison of sequences or restriction fragment length polymorphism (RFLP) frequency tables (e.g. Cann *et al.*, 1987).

Analysis of haplotype relatedness by phylogenetic methods can give startling insights into human evolution. Cann *et al.* (1987) demonstrated, by phylogenetic analysis of the mitochondrial genome, that all modern humans could be placed on one phylogenetic tree rooted in Africa 200,000 years ago. An assumption in this analysis is, however, that the rates of mutation (and hence presumably mutation fixation) have remained constant in evolutionary history (Kocher *et al.*, 1989). Analysis of more recent human radiations from the timing of human population expansions (e.g. Sherry *et al.*, 1994) to the determination of ethnicity progression (for review see Torroni and Wallace, 1994) have all been possible with phylogenetic analysis using mtDNA.

Mitochondrial DNA diseases

As mtDNA has few intergenic spacers and mutations occur at random positions, the coding region accumulates many mutations. Whilst relaxed codon usage limits the damage these mutations may do, many are known to be deleterious. All 13 polypeptides encoded by mtDNA are integral to OXPHOS, therefore deleterious mutations invariably cause a decrease in ATP production. In addition to the 13 polypeptides, the 22 tRNAs and 2 rRNAs can also contain deleterious mutations. As no nuclear encoded tRNAs are imported into mitochondria from the cytoplasm, mutations in the tRNA genes can have severe effects on mitochondrial protein synthesis. An increasing number of diseases have been attributed primarily to mtDNA mutations (for review see Brown and Wallace, 1994). Many of these diseases share similar characteristics of inheritance and penetrance. In germ and somatic cells, there are many mitochondria, frequently interconnected, but all containing many copies of mtDNA. All mtDNA mutations initially exist as a single copy of mtDNA in a cell of an individual. These mutations, providing they do not affect replication, are reproduced when the mtDNA replicates. During mitosis, the mutation containing mtDNA segregates either randomly or selectively to daughter cells. Cells, tissues, or individuals that have more than one type of mtDNA are said to be heteroplasmic. Cells can withstand a certain amount of deleteriously mutant mtDNA before they are phenotypically affected. The point at which the abnormal phenotype becomes detectable is called the threshold, and is dependent on the balance of ATP requirement and optimal ATP production. The threshold can vary between cells and tissues, and depends on the type of mutation and whether there are other changes that can act as secondary mutations. Deleterious mutations (with exceptions) are seldom passed on as homoplasmic changes as these have a negative effect on fitness. They can, however, be

inherited heteroplasmically for a number of generations. All mutations go through a state of heteroplasmy. Selectively advantageous or neutral mutations tend to become fixed to homoplasmy or get lost rapidly, with the rate often exceeding that expected by chance alone (Hauswirth and Laipis, 1982). In some instances heteroplasmic changes are fixed to homoplasmy in as little as one generation, but in others they are maintained in a heteroplasmic state, for many generations (Howell *et al.*, 1996) (see Chapter 4). The mechanisms for maintenance of heteroplasmy are poorly understood, but may involve selective amplification of rare 'alleles' within developing oocytes, a so called 'bottleneck' effect (Poulton, 1995). Given that all mtDNA mutations must go through a state of heteroplasmy, and that the mtDNA mutation rate is high, one would expect that heteroplasmy is common. However, until recently, little heteroplasmy has been found.

Heteroplasmy (discussed in Chapter 4), and indeed mtDNA diseases, can be divided into two groups according to the nature of the mutations. Those caused by nucleotide substitutions and those caused by insertions and deletions. Nucleotide substitutions can occur anywhere in the mtDNA, but when causing disease, the changes usually occur in the conserved nucleotides in the protein coding region or tRNA or rRNA coding regions. For example Leber's hereditary optic neuropathy (LHON), a maternally inherited form of adult acute-onset vision loss, can be caused by a number of nucleotide substitutions. Sixteen LHON mutations have been characterised to date, 11 of which occur in the mitochondrial ND genes (from OXPHOS complex 1). The most common of these changes is a G $\rightarrow$ A transition at nucleotide (nt) 11778 (Wallace *et al.*, 1988) which is present in over 50% of LHON patients. It changes a conserved arginine to a histidine residue in ND4. In addition to primary mutations there are a number of characterised secondary mutations that are mildly deleterious, accumulate over time, and predispose an individual to LHON. Such mutations require additional factors to cause LHON, and therefore are generally associated with late onset of the disease (Howell *et al.*, 1993). Primary LHON mutations are frequently homoplasmic. Between LHON lineages, incomplete penetrance and the presence or absence of secondary mutations creates a complex population picture of mitochondrial diseases.

Like nucleotide substitutions in protein coding genes, those in tRNA and rRNA coding regions are also frequently heteroplasmic. An A $\rightarrow$ G transition in $tRNA^{Lys}$ at nt 8344 gives rise to myoclonic epilepsy with ragged red fibres (MERFF) (Silvestri *et al.*, 1993). This maternally inherited disease, which causes uncontrolled jerking (myoclonic epilepsy) and

myopathies is associated with a decrease in the synthesis of complex I and IV mitochondrial proteins. Likewise an A → G at nt 3243 within $tRNA^{Leu(UUR)}$ causes mitochondrial encephalopathy, lactic acidosis, and stroke like episodes (MELAS) (Penn *et al.*, 1992). Both MERFF and MELAS are generally heteroplasmic, with the severity of the phenotype being directly dependent on the percentage of mutant DNA in cells, tissues, and individuals; and the age of the patient. The age of the patient is thought to be important because older people do not have as efficient OXPHOS as younger people, possibly because of a lifetimes' accumulation of secondary mutations in post-mitotic tissues. Therefore, a small proportion of mutant DNA may create a disease phenotype in an older person, while a larger proportion of the same mutant is needed to produce the disease phenotype in a younger person (Shoffner *et al.*, 1990). Mitochondrial DNA diseases caused by nucleotide substitutions typically require a tissue to have greater than 80% mutant DNA for a disease phenotype to manifest itself. As mentioned, this threshold is both age and tissue specific. The earliest affected tissues (those with the lowest threshold and highest ATP demand) have been characterised as brain, kidney, heart, and liver (in this order) (Brown and Wallace, 1994). In addition, type I muscle fibres are readily affected, and if present in the pancreatic islet cells, mutations that cause a reduction in OXPHOS inhibit insulin secretion leading to late onset diabetes (type II).

The second type of mtDNA mutation, insertions and deletions, can also cause disease phenotypes. Diseases caused by deletions range from the mild chronic progressive external opthalmoplegia (CPEO) to the severe Kearns-Sayre syndrome (KSS). Patients with these disorders are clinically variable, although most have ophthalmoplegia (paralysis of the eye muscles), ptosis (droopy eye lids), and mitochondrial myopathy. Patients with KSS can also have retinitis pigmentosa, hearing loss, heat conduction defects, ataxia, and dementia (Wallace, 1992). Although over 100 deletions are associated with these and other conditions, the deletions share many common properties. Firstly, they are rarely maternally inherited, arising spontaneously within an individual. There are examples of autosomal dominant deletion inheritance, where the deletions are not inherited *per se*, but the propensity for deletion occurrence is (Suomalainen *et al.*, 1995). Secondly, deletions commonly occur between direct repeats, the most common deletion occurring between the largest direct repeat, suggesting that a common mutagenic mechanism is slip-replication (eg. Johns *et al.*, 1989; Shoffner *et al.*, 1989). Thirdly, most deletions remove at least one tRNA, thus affecting many OXPHOS complexes (Wallace, 1992). Fourthly, deletions tend to spare the heavy and light strand origins of replication, allowing replication to proceed on the shortened molecule.

Because of the decreased length of the molecule, replication is faster than for normal length molecules, and therefore the deleted molecule increases in proportion temporally in affected individuals. Conversely, mtDNA insertions causing Pearson's syndrome incorporate both replication origins. The doubling of replication origins in inserted molecules presumably leads to an increase in replication and could explain the selective enrichment of these larger molecules characteristic of this syndrome (Brown and Wallace, 1994).

Neutral mtDNA mutations

In addition to the above insertion and deletion events, all of which are heteroplasmic and deleterious, there are a number of presumably silent insertion and deletion events characterised. A 9 bp deletion within an intergenic spacer between COII and $tRNA^{Lys}$ is commonly found in African, Polynesian and some American Indian populations (eg. Sykes *et al.*, 1995). In the same intergenic spacer, a simple 4 bp homopolymeric insertion has been seen presumably caused by replication slippage (discussed in Chapter 3). In the D-loop such homopolymeric insertions are frequently seen in heteroplasmy, with variable lengths of tracts being added. All of these insertions and deletions are, however, assumed to be neutral, given that they cause no phenotypic effects, even in homoplasmy.

The assignment of nucleotide sequence changes as either polymorphisms or deleterious mutations is difficult. Riordan-Eva and Harding (1995) used a number of criteria to distinguish between the two. Firstly, deleterious mutations should be found specifically in individuals with the disease and not in normal individuals. Secondly, the mutations should change a highly conserved nucleotide or amino acid. Thirdly, the mutations should not be found with other disease causing mutations. Fourthly, the mutations should be heteroplasmic, with a large amount of mutant DNA being required before the mutant phenotype is displayed. In addition they defined a need for maternal inheritance. This and other points, are in dispute. Many mutations such as deletions arise spontaneously or may be inherited in an autosomal dominant fashion. Some LHON mutations are homoplasmic but fit all of the other criteria. Many secondary mutations can be found in normal individuals but only have phenotypic effects in the presence of other mutations. Thus the distinction between deleterious mutation and polymorphic sequence variant is often blurred. Recent evidence suggests that even seemingly silent, haplotype-defining, polymorphisms may be markers for mtDNA disease susceptibility (e.g. Hofmann *et al.*, 1997).

Mitochondrial DNA mutations, selection, and fertility

The predominantly uniparental maternal inheritance of mtDNA creates interesting selection issues. Any mutation affecting males is effectively neutral on a population level because:

> "Maternally transmitted mutations that enhance male fitness tend not to increase deterministically within populations, since their beneficial effects on males are not rewarded by increased representation among later generation progeny; and female-transmitted alleles with male-limited deleterious effects may be retained simply because of relaxed direct selection against their loss" Avise (1991).

In fact a germ-line mitochondrial mutation with severe effects on males, but only mild effects on females, could increase to relatively high population frequencies because natural selection of mitochondria occurs only in females.

At least one maternally inherited mitochondrial disease affects males more severely than females. Between 75 and 85% of individuals affected by LHON are male (Brown and Wallace, 1994). As mentioned, 16 distinct base-pair substitutions in germline mtDNA have been directly associated with this disease. Linkage analysis initially suggested that a deleterious X-linked mutation in combination with defective mitochondria may be the cause of this sex bias (Vilkki *et al.*, 1991). Juvonen *et al.* (1996) and others found no further evidence for this, and indeed an analysis of X-inactivation by Oostra *et al.* (1996) further argues against an X-linked effect to this sex bias. Males may just be more susceptible to the mitochondrial defects. Wallace (1992) suggests that men may have a higher optic nerve metabolic rate than women, and therefore a lower threshold. Theoretically male-female dimorphism is possible in a number of tissues, including the heart (Frank and Hurst, 1996).

Males germ cells (sperm) may be preferentially affected by deleterious mtDNA mutations. Sperm have around 50 copies of mtDNA held in 13 mitochondria, spirally-arranged in the sperm midpiece (Folgerø *et al.*, 1993). There is tight coupling between mitochondrial energy production and sperm motility, with almost all energy produced by

the mitochondria being used to drive the flagellum and propel the sperm (e.g. Kao *et al.*, 1995). It must be noted, however, that much of the ATP used in flagellar action is derived from glycolysis, not OXPHOS (Ford and Harrison, 1981). Additionally, within fertile men, the level of seminal ATP is not predictive of fertilising potential[3] (Cummins *et al.*, 1994). Reduced power output translates directly into reduced motility and probably reduced fertilising potential. The effect that deleterious mtDNA mutations, have on sperm and ultimately fertilising potential through ATP reduction, however, has to be weighed against the fitness effects in females. If the same effects are seen in females in other tissues, then these changes will be rapidly selected against and lost, and the ultimate effect on population male fertility will be low.

Human male fertilising potential

Fifteen percent of couples have problems conceiving. Of these, 30% are due to a male cause, with another 20% being related to incompatibility of both partners. Infertility (or subfertility) is difficult to define as it is related to a couple's mutual potential to reproduce. The World Health Organisation (WHO, 1992) defines infertility as an inability to conceive after 12 months of regular intercourse. Lack of male fertilising potential can be attributed to a variety of factors. Based on WHO guidelines (WHO, 1992) a male is considered to have normal fertilising potential if his ejaculate sperm count is greater than 20×10^6 sperm/ml of semen, at least 50% of which are motile, and at least 50% having normal morphology. Values below these yield the clinical conditions oligozoospermia, asthenozoospermia, and teratozoospermia respectively.

Low male fertilising potential can be caused by a large number of factors that can be divided into five categories. Obstruction of excurrent testicular or ejaculatory ducts accounts for 7 to 15% of low male fertilising potential. Testicular lesions causing spermatogenic failure account for a further 30%. These lesions may include testicular trauma, varicocele, genital tract infection or Y-chromosome deletions. Erectile or ejaculatory problems, whilst rare, can also cause fertilising potential decrease; as can endocrine disorders, such as hypogonadotropic hypogonadism. The most frequent cause, however, is termed idiopathic sperm dysfunction (i.e. dysfunction with no known

[3] No individual has inherent fertility, fertility is only proven after a child is conceived. Therefore the term 'fertilising potential' is used in this study to describe the potential to fertilise an egg.

aetiology) (Cummins *et al.*, 1994). Causes for this may include antisperm antibodies, drug or alcohol use, smoking, environmental factors, and undetermined genetic factors.

Fertilising an egg is a chance event. The ability to fertilise an egg *in vivo* is enhanced with larger numbers of highly motile sperm with normal morphology. Conversely, the chances are reduced with fewer active sperm. In this sense, an ejaculate of seminal fluid can be considered as a tissue. A 20% decrease in motility could decrease an individual's fertilising potential by an equal amount, possibly rendering the affected couple effectively infertile. The very tight coupling between mitochondrial energy production and motility, and between motility and fertilising potential suggests that if deleterious mtDNA mutations were present within sperm they could potentially cause a reduction in fertilising potential (Frank and Hurst, 1996). This has been demonstrated in a number of studies (e.g. Folgero *et al.*, 1993; Kao *et al.*, 1995).

There are two types of deleterious mtDNA mutations that have been shown to possibly affect male fertilising potential. Firstly, mtDNA deletions within the testes have been correlated with spermatogenic failure, resulting in severe oligozoospermia. Cummins *et al.* (1993) first showed that spermatogenic failure in a 36 year-old man might be caused by premature testicular aging as a result of a 4977 bp deletion in the mitochondrial genome, a deletion which was first described in normal heart muscle and brain of older people (Cortopassi *et al.*, 1992). Kao *et al.* (1995) then showed that such mutations were relatively common in oligozoospermic men, but were also found in normal men. Significantly, they showed that the proportion of the mutation present in the semen correlated very well with the sperm motility ($r = -0.98$, Kao *et al.*, (1995); Fig. 3 therein), supporting the tight coupling between mitochondrial energy production and motility. Others, however, have not found such a good correlation (St John *et al.*, 1997). Lee *et al.* (1994) could not detect the mutation in the testes until the age of 60 years. The mechanisms causing these deletions are unknown. Cummins *et al.* (1994) speculates that the deletions are created in the testes by premature testicular aging due to an ischaemic injury. Recent evidence that mutations can be induced in rat testes by ischaemic injury supports this hypothesis (A.M. Jequier, pers. comm.). The presence of the mtDNA mutation within testes could explain the lack of mature sperm being produced (oligozoospermia) whilst the mutation presence in sperm may explain a reduction in motility (asthenozoospermia).

Mitochondrial DNA point mutations have also been found in cells within semen. Folgerø *et al.* (1993) demonstrated that a 50 year old male patient with MERRF with an A → G transition at nt 3243 in the mitochondrial $tRNA^{Leu(UUR)}$ gene also had reduced sperm motility. Although the proportion of mutant DNA within the testes was not stated, blood platelet DNA from other members of the proband's family had between 13 and 59% mutant DNA (Folgerø *et al.*, 1995). The effect on fertility was noted by seminal substrate therapy. Different energy substrates can enter the electron transport chain (ETC) at different positions. For example, glucose and lactate are both metabolised to pyruvate which enters the ETC in complex 1. Succinate, however, bypasses complex 1, and increases ATP production at complex III and possibly complex IV. In this way, if complex 1 is affected by mtDNA mutations, ATP production can be increased with the addition of succinate. Incubation of semen in succinate selectively increased the motility of sperm from the proband whilst having limited effect on semen of control samples (Folgerø *et al.*, 1993). Therefore, it was concluded that the motility of sperm was directly affected by this mtDNA mutation. Contrary to Folgerø *et al*'s report, Huang *et al.* (1994) failed to detect changes in the sperm motility of a patient with MELAS who had the same tRNA mutation. They estimated that 30% of amplified mtDNA from cells within semen contained the mutation, and found that sperm motility was not different from control samples. They did not, however, demonstrate that the mutation was present in sperm *per se*, only in seminal cells, and did not attempt to increase motility by substrate therapy. Semen is a mixture of different cell types, frequently containing leucocytes, immature spermatids, and epithelial cells from spermatogenic ducts. All of these will contribute mtDNA to a seminal preparation. These two conflicting reports, however, show that it is still unclear whether point mutations in sperm mtDNA can affect sperm function.

All of the above sperm mtDNA mutations have known deleterious effects in other tissues. As suggested earlier, a decrease by 20% of sperm motility may cause a pathological reduction in fertilising potential. Given the coupling between energy production and motility, this could be caused by a 20% proportion of deleteriously mutant mtDNA with the sperm. No other tissue is likely to be affected by a proportion this low in males, and indeed no tissue in females will be affected at all. This then begs the question: 'Could the threshold of seminal fluid, to withstand the effects of mtDNA mutations, be very low ?' Given that there is no selection against mtDNA mutations in males, these mutations may be relatively common, as long as the fitness effect when the mutations are in females is low (as would be the case if only a small amount of mutant

DNA were present). The incidence of mtDNA defects causing a decrease in male fertilising potential is presently unknown.

This study was instigated to analyse some of the above phenomena. When examining the effect mtDNA mutations have on male fertilising potential, there are two possible approaches to the problem. The first method involves recruiting subjects with known mitochondrial disease and investigating their seminal parameters (akin to the work of Folgerø *et al.*, 1993). However, there are a number of problems with this approach. Firstly, the number of patients in New Zealand with defined mitochondrial disease is small. Secondly, patients with defined mitochondrial disease are often severely affected, frequently not living to reproductive age. And thirdly, an analysis of the fertilising potential in patients with mitochondrial disease is not predictive of the incidence of reduced fertilising potential being caused by mtDNA mutations in the general population. Because of these reasons a different approach was taken in this study. This involved recruiting patients in an infertility clinic and screening seminal samples for mtDNA variation. Like the earlier approach, this has shortcomings. The mitochondrial genome is large and point mutations are small. Because of this only three coding genes are screened for mutations in the present study. As the work of Folgerø *et al.* (1993) showed substrate therapy can ameliorate the effects of some mitochondrial defects. Given that these substrates (such as succinate) may be naturally present in semen, we hypothesised that mutations in complexes that cannot be by-passed by natural substrates may show greater phenotypic effects than mutations in complexes that can. Therefore in this study, the ATPase 6 and 8 genes from the F_0F_1 ATP synthase (unofficially called complex V) and the COII gene from complex IV were screened for mutations. Recent studies carried out in our laboratory on complex I genes found no heteroplasmy associated with fertilising potential (S.J. O'Carrol, unpublished MSc thesis, University of Canterbury, 1996). The work reported here, therefore, is part of an on going project in our laboratory, with other genes such as Cyt b and COIII being studied.

The aims of the present study are:

1. To instigate a (biased) population study of mtDNA variability in the semen samples from men presenting samples to an andrology clinic in Christchurch, New Zealand.

2. To screen mtDNA protein coding genes ATPase 6, ATPase 8 and COII for sequence variability.

3. To investigate correlations between the presence or absence of mtDNA mutations and seminal quality.

4. To characterise new mtDNA heteroplasmy within human sperm for the analysis of :

    a. Fertilising potential

    b. Mitochondrial DNA variation

Chapter 2

ATPASE 6 AND 8 GENE MUTATIONS

Introduction

The ATPase amplicon

The mitochondrial F_0F_1 ATP synthase utilises an electrochemical gradient, generated by the electron transport chain, to synthesise ATP from ADP + Pi. Called complex V, this ATP synthase (incorrectly but commonly called an ATPase) along with the electron transport chain complexes (I, II, III, IV) and adenine nucleotide translocator constitute the mitochondrial energy generating pathway, OXPHOS. Intact ATP synthase was first isolated from bovine heart mitochondria in 1974 (Hatefi, 1993). Mammalian mitochondrial F_0F_1 ATP synthase is composed of at least 13 polypeptides, two of which (ATPase 6 and ATPase 8) are encoded by the mitochondrial genome and synthesised in the mitochondria. ATPase 8 is also commonly called A6L. The coding regions of ATPase 6 and 8 overlap by 46 nucleotides (8537 - 8572, numbering from the Cambridge sequence) at the 5' end of ATPase 6. ATPase 6 and 8 have different reading frames (Anderson *et al.*, 1981). Both ATPase 6 and ATPase 8 are found within the F_0 subunit of F_0F_1 ATP synthase. This subunit acts as a membrane anchor and proton pore, facilitating proton movement through the inner mitochondrial membrane across an electrochemical gradient. Within the F_0 pore, it is thought that ATPase 8 is in a 1 : 1 molar ratio with the pore itself (Hekman *et al.*, 1991). It is anticipated that ATPase 6 is in the same molar ratio, consistent with ATPase 6 and ATPase 8 both being transcribed on the same mRNA. In addition to the two mitochondrial subunits, five additional nuclear encoded subunits make up the F_0 pore (Hatefi, 1993).

The structure of ATPase 8 is presently unknown. Hekman *et al.* (1991) demonstrated that parts of ATPase 8, including its C-terminus, were exposed on the matrix side of the inner mitochondrial membrane in bovines. Consistent with the lack of known structure is the lack of known function, and the lack of pathogenic mutations affecting structure and function.

Structural models of the human ATPase 6 protein are based on models of the homologous *Escherichia coli* F_0 ATP synthase subunit *a* protein. Multiple mutational studies have been performed on the *E. coli* protein to elucidate its secondary structure and residues

important for its function in the F_0F_1 ATP synthase. Cox *et al.* (1986) proposed a five transmembrane helix (a1 - a5) secondary structure model for both the *E. coli* and human proteins (Fig. 2.1). Homology studies, site-directed mutagenesis studies and naturally occurring mutations have provided evidence that conserved residues in helix a4 are essential in allowing the F_0 subunit of the F_0F_1 ATP synthase to function as a proton pore (for review see Senior, 1990).

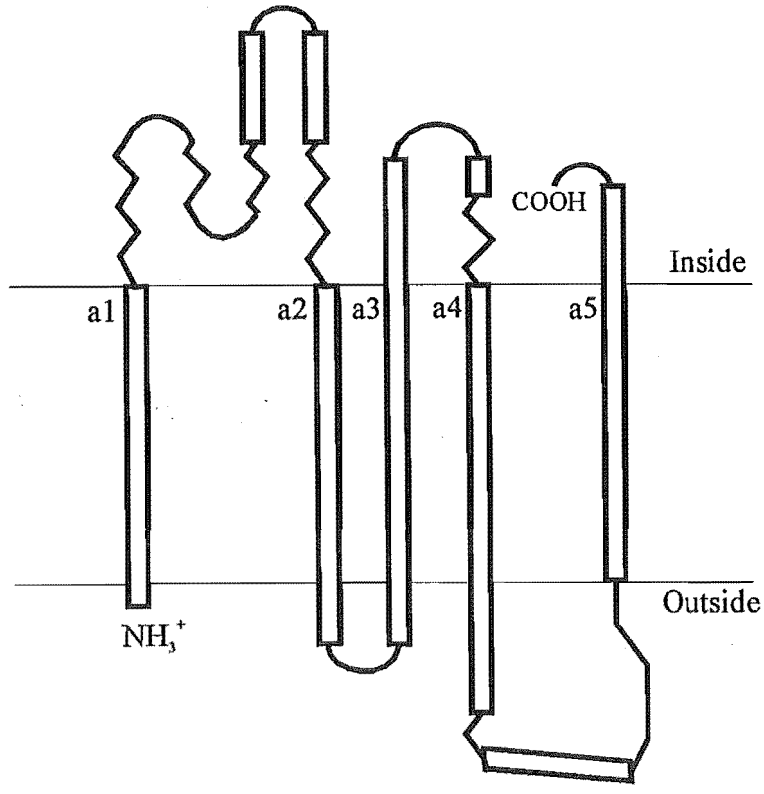


Figure 2.1. Proposed structure of the human ATPase 6 protein (modified from Cox *et al.*, 1986). Membrane spanning helicies are labelled a1 - a5. Inside and outside refer to the orientation within the inner mitochondrial membrane.

Five deleterious mutations have been identified in the ATPase 6 gene. The most common of these is a T → G transversion at nt 8993. It changes a highly conserved leucine to an arginine at residue 156 in helix a4 of the ATPase 6 protein (Holt *et al.*, 1990). A T → C transition at the same position has been seen, changing the Leu^{156} to a proline (de Vries *et al.*, 1993). These mutations cause diseases with varying severity ranging from neurogenic muscle weakness, ataxia and retinitis pigmentosa (NARP) (Holt *et al.*, 1990) to subacute necrotizing encephalopathy (Leigh's disease) (Tatuch *et al.*, 1992). They have been found both in homoplasmy and heteroplasmy but require high proportions of mutant DNA for phenotypic

symptoms. Leu^{156} is considered to be an integral residue for proton translocation across the inner mitochondrial membrane as demonstrated in *E. coli* and yeast studies (de Vries *et al.*, 1993). A T → C transition at nt 8851 has been described in a patient with bilateral striatal necrosis (similar to Leigh's disease) (de Meirleir *et al.*, 1994). This mutation changes tryptophan residue 109 to an arginine in helix a3. Similarly, a T → C transition at nt 9176 was found in two brothers, also with bilateral striatal necrosis (Thyagarajan *et al.*, 1995). This alters the highly conserved leucine residue 217 to a proline. The mutation only manifested phenotypically when the proportion of mutant DNA was greater than 98%. Leu^{217} is in the C-terminus of ATPase 6, a region which has been proposed to be involved in proton translocation (Hartzog and Cain, 1993). A T → C transition at nt 9101 has been associated with LHON (see Chapter 1) (Lamminen *et al.*, 1995). However, the affected proband and three unaffected maternal relatives all had this change in homoplasmy. Additionally, the amino acid residue affected (Ile^{192}) within helix a5 is very weakly conserved even in vertebrates. Based on these last two points the T → C at nt 9101 may at best be a secondary mutation.

The incidence of mtDNA defects causing a decrease in male fertilising potential is presently unknown. Since mutations in the ATPase 6 and 8 genes are known to affect the phenotypes of individuals carrying the mutations and the protein products of these genes are integral products of complex 5, the penultimate OXPHOS complex, and as such can not be bypassed by natural energy substrates (see Chapter 1), a mutation in these genes may affect sperm function.

The aims of the present study were to assess the variability in the ATPase 6 and 8 genes and identify mutations in these genes in a Christchurch male population and to determine if these mutations can be related to the fertility status of the individuals carrying the mutations.

In this study, mutation screening (described below) was facilitated by polymerase chain reaction (PCR) amplification of the ATPase 6 and 8 genes in an ATPase amplicon. This amplicon spans not only these genes but regions of COII and COIII and also $tRNA^{Lys}$. COII and $tRNA^{Lys}$ are covered in the following chapter. As shown in the results, no changes were found in COIII and so this will not be mentioned further. To detect sequence variability in the ATPase amplicon single-strand conformation polymorphism (SSCP) analysis

was used (Orita *et al.*, 1989). SSCP relies on the fact that when single stranded DNA is cooled quickly it folds up on itself forming a secondary structure. When separated through non-denaturing polyacrylamide gels, these secondary structures have a mobility dependent on both size and shape. A difference of one base pair between two otherwise identical DNA fragments frequently changes the shape of one and hence creates a mobility shift with respect to the other.

Materials and Methods

Sample collection

Subjects were recruited from couples who presented for semen analysis at Christchurch Women's Hospital between January 1994 and the January 1997. Included in this group are couples who have male factor infertility, female factor infertility, or infertility of unknown aetiology. Additionally, semen from some donors involved in the donor insemination program and samples produced prior to long term storage were included. Ethical approval was obtained from the Southern Regional Health Authority Ethics Committee (Canterbury). The donors gave consent for the residue of their semen samples (after routine semen analysis) to be used for research. Semen samples were produced by masturbation after 3 - 5 days of abstinence. The semen samples were analysed at the Andrology Laboratory, Christchurch Women's Hospital, according to standard criteria (WHO, 1992). Whole semen samples were transferred frozen to the Zoology Department, University of Canterbury, thawed on ice and re-aliquoted to 100 µl aliquots. These aliquots were snap frozen to -80°C for long term storage. For confidentiality reasons, sample information used in this study was restricted to non-identifying sample codes with sperm counts, sperm motility, seminal volume and year of birth of the donor. All screening in this study was non-biased. Samples were sequentially chosen from the pool irrespective of count or motility. Within the sample pool there are cases of repeated samples from donors.

For ATPase amplicon screening, 210 semen samples were analysed with a mean count of 69 x 10^6 sperm/ml (± 78), a mean sperm motility of 48% (± 20) and a mean donor age (at time of donation) of 36. Full details of the samples used in this screening are included in Appendix 1.

Cell DNA isolation

The preferential purification of mitochondrial DNA generally requires large volumes of starting tissue. Frequently, after routine semen analysis, the volume of semen obtained for this study was under 100 µl. Therefore in seeking a DNA extraction protocol the following points were considered. (1) Sperm have relatively less mtDNA than other tissue types, (2) the volume of sample was a limiting factor, and (3) if possible all seminal mtDNA had to be available for PCR (see Chapter 4). For the above reasons a modified buccal cell alkaline lysis protocol (Lench *et al.*, 1988) was used for the extraction of mtDNA from semen. This method requires little manipulation, can be performed in a single tube (ensuring no loss of DNA by repeated extractions), and is ideal for small volumes of starting material.

Semen samples were thawed on ice. Cells were pelleted from 5 µl aliquots by centrifuging at 4,500 x g, 4°C for 10 min (Jouan MR 14.11 centrifuge). The cell pellet was washed with 400 µl of 100 mM NaCl, 10 mM EDTA (pH 8.0), then re-centrifuged as before. The cell pellet was resuspended in 40 µl of 50 mM NaOH by vortexing, then boiled for 15 min and neutralised with 10 µl of 1 M Tris-HCl (pH 8.0). Ten-microlitre aliquots were frozen at -20°C until further use. The impurity of this DNA preparation meant that no quantification of DNA could be made, suffice as to say that between 2.5 and 5 µl of prepared DNA was sufficient for PCR amplification irrespective of the cell count of the original semen sample.

In one instance mtDNA was isolated from whole blood by first lysing red cells in a 50 µl aliquot of whole blood with 1 ml of 0.14 M NH_4Cl. This was shaken gently for 10 min before pelleting white cells in a bench top microfuge at 550 x g for 2 min. The cells were washed with 1 ml of 0.14 M NH_4Cl and pelleted again as before. The pellet was resuspended in 400 µl of 50 mM NaOH and boiled for 15 min as above then neutralised with 100 µl of 1 M Tris (pH 8.0). Five-microlitre aliquots provided sufficient template for PCR amplification.

Polymerase chain reaction

ATPase amplicon primers were chosen using the OSP computer program (Hillier and Green, 1991) in such a way that they did not fall on common mtDNA polymorphic

sites, did not anneal to themselves or each other with high affinity and had melting points above 50°C for high specificity binding. The primers chosen were HMTL817 (5' CAATGCTCTGAAATCTGTGG) and HMTH934 (5' tagtatgaggagcgttatgg). These primers amplify a 1174 bp ATPase amplicon between nt 8186 and nt 9341 (numbering from the Cambridge sequence). The ATPase amplicon encompasses genes for ATPase 6, ATPase 8, $tRNA^{Lys}$, 101 bp of COII and 134 bp of COIII. PCR conditions for the ATPase amplicon were optimised for $MgCl_2$ concentration and annealing temperature.

For the analysis of the ATPase amplicon, a 5 µl aliquot of total cell DNA was subjected to 30 cycles of touchdown PCR in a 25 µl reaction. The 25 µl reaction mixtures contained 20 pmol of each primer, 5 nmol of each dNTP, 2.5 µl of 10 x reaction buffer (500 mM KCl, 100 mM Tris-HCl (pH 9.0)), 2 mM $MgCl_2$ and 0.75 units of *Taq* polymerase (Promega or Boehringer Mannheim). The cycling parameters were an initial five cycles of 94°C/30 sec, 60°C/30 sec, and 72°C/1 min 20 sec followed by a further five cycles of 94°C/30 sec, 56°C/30 sec, and 72°C/1 min 20 sec. This was followed by 20 cycles of 94°C/30 sec, 54°C/30 sec, and 72°C/1 min 20 sec. Prior to initiation of the PCR, reaction mixes were equilibrated at 80°C for between 2 and 5 min. This equilibration step was found to be essential for highly specific, and efficient product formation. Thermal cycling was performed in two different PCR cyclers (Perkin Elmer 480 and MJ Research PTC 100) with no product variability being seen between these machines.

Mutation Detection

The 1174 bp ATPase amplicon PCR product from all 210 semen samples was cleaved to completion by *Taq*1 restriction endonuclease prior to SSCP analysis. Ten microlitres of PCR product (100 - 200 ng) was cleaved with 2 units *Taq*1 (Boehringer Mannheim) at 65°C to yield fragments of 789 bp and 384 bp. No star or partial activity was observed with *Taq*1 digestion. For SSCP analysis, 3 µl of digested DNA (15 - 30 ng) was heat denatured with 8 µl of formamide gel loading buffer (95% formamide (v/v), 20 mM EDTA, 0.05% Bromophenol blue (w/v), 0.05% xylene cyanol FF (w/v) (Poon *et al.*, 1993)) for 5 min at 95°C before being plunged onto ice and left for at least 5 min. Samples were

electrophoresed on 5% non-denaturing polyacrylamide gels (49% acrylamide : 1% bis-acrylamide) (Appendix 2) for 8 hours (4°C, 200 V). Gels were pre-run to equilibrate in 45 mM Tris-HCl, 45 mM Orthoboric acid and 1 mM EDTA (0.5 x TBE). Gels were fixed and silver stained using a modification of the protocol of Bassam *et al.* (1991) (Appendix 2). Band densities, where appropriate, were determined from dried gels using a hand held scanning densitometer and Biomed Image Analysis software (Advanced American Biotechnology, Fullerton, CA).

Undenatured DNA, when electrophoresed on SSCP gels gave single bands that did not interfere with scoring of denatured bands. To assign fragment sizes to denatured bands, both of the ATPase amplicon *Taq*1 fragments were isolated from agarose gels individually by either a 'freeze and squeeze' protocol (Towner, 1993) or using a commercial gel extraction kit (BioRad or Boehringer Mannheim) and were electrophoresed as denatured and undenatured forms on SSCP gels.

Where changes in SSCP pattern were observed during screening, the samples were analysed again to confirm the change in pattern. The ATPase amplicons from samples with consistently different patterns were then cleaved with different restriction endonucleases and re-analysed by SSCP analysis. The DNA fragments carrying the nucleotide changes were identified by comparing the mobilities of the fragment bands with those of the controls.

Each of the two single strands (ssDNA) of a fragment of double stranded DNA (dsDNA), when denatured and allowed to form a secondary structure, will fold differently. Because of this, each fragment, a single band when undenatured, will move as two bands in an SSCP gel. On occasion, one of the single strands may form more than one stable secondary conformation, creating three or more bands representing the original dsDNA fragment. A homoplasmic or polymorphic change can be recognised by one or both of the ssDNA bands having altered gel mobility when compared with a control sample. While a heteroplasmic change can be identified when one or more new bands are seen in addition to the bands which are present in the control sample (Fig. 2.2).

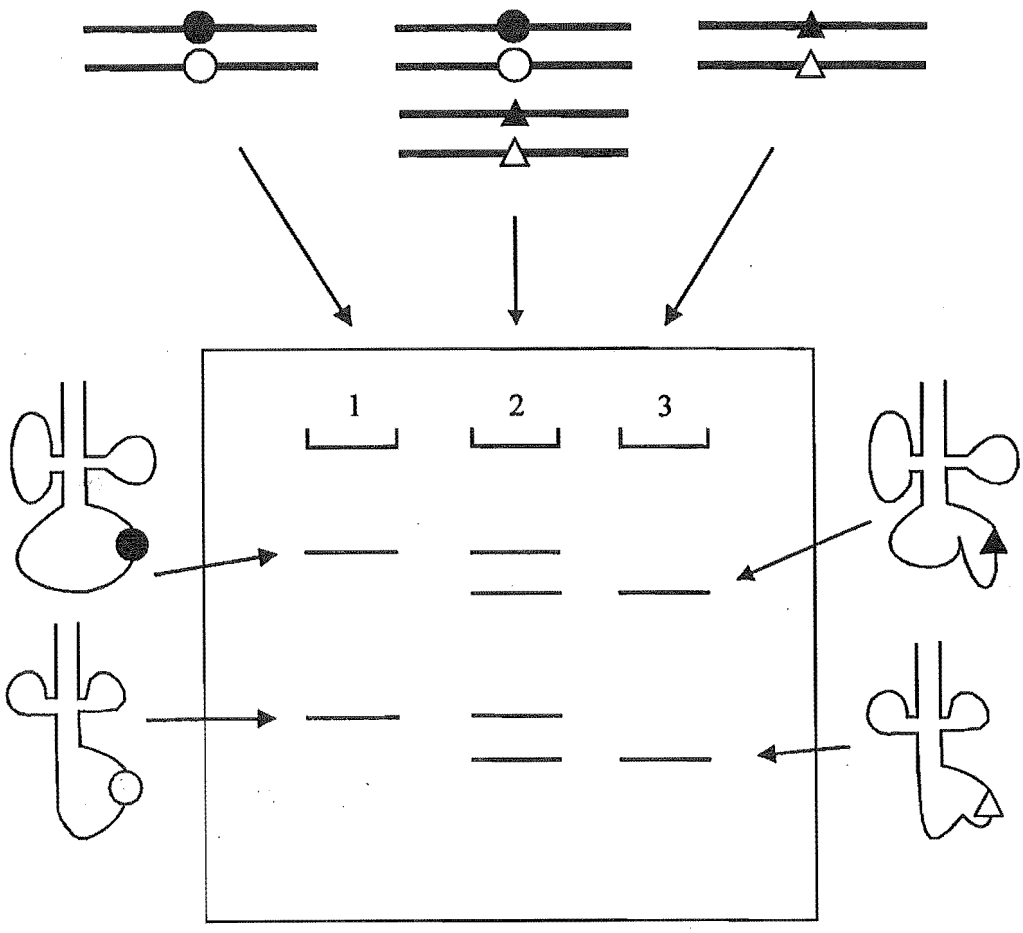


Figure 2.2. Diagrammatical representation of SSCP analysis. Circles represent normal bases with triangles being the mutated counterparts at the same position. Filled and empty shapes represent bases on different strands of the same fragment. When resolved, homoplasmic normal samples give two bands per fragment as shown in lane 1. Heteroplasmic samples (lane 2) have both normal bands (as lane 1) and one or two additional bands representing the mutant strands. Lane 3 represents a homoplasmic mutant where one or both of the bands runs in a different position to the normal (lane 1).

Single cell analysis and sequencing

Heteroplasmy, as seen in SSCP gels, can be analysed in a number of ways. (1) If the mutant DNA is greater than 20% of the total DNA (Smith *et al.*, 1992), direct sequencing of the mixture will often resolve the mutation. If the sequencing creates any other artefacts, however, this will not work as the sought after variation will also appear as an artefactual band or peak. (2) Mutant bands can be extracted directly from SSCP gels and amplified to enrich the mutant component for sequencing (Suzuki *et al.*, 1991; Sherratt *et al.*, 1996). This can be done only if the DNA is stained with ethidium bromide (which only weakly stains single stranded DNA) or is isotopically labelled. Silver stained DNA can not be re-amplified. (3) The heteroplasmic region of interest can be cloned into plasmids; the resulting clones will be either mutant or normal (see Chapters 3 and 4). (4) Finally, and potentially most

informatively, single cells can be isolated, and the DNA amplified and analysed. As different cells have different proportions of mutant mtDNA in heteroplasmy, mutant enriched cells can be analysed and the mutant mtDNA sequenced. This procedure was chosen for the analysis of an ATPase amplicon heteroplasmy in two samples from one individual (see results).

Single cells were aspirated from a 1,000 times diluted semen sample under oil using micromanipulation techniques. Briefly, 1 mm O.D. glass needles (Clark Electromedical Instruments) were pulled to a 2 μm O.D tip using a Narishige microelectrode puller. Tips were ground to a 45° bevel with an inner hole diameter of 5 μm using a Narishige EG-7 microgrinder. Needles were sterilised by baking at 150°C for 4 hours and then exposing to U.V. light for 20 min. Cells were resuspended in PBS (120 mM NaCl, 16 mM Na_2HPO_4, 4 mM NaH_2PO_4 (pH 7.3)) at 10,000 cells per ml and 20 μl was transferred to the base of a Lux 5221 tissue culture dish. This drop was covered in sterile mineral oil (Sigma) to prevent optical edge effects and cells were visualised under a Leitz Diavert inverted microscope. Single cells were aspirated into the bevelled tip of a needle held in an M3301 Micromanipulator (World Precision Instruments). The fine control for cell pick-up was provided by holding pressure from a PV830 Pneumatic Picopump (World Precision Instruments) attached to a N_2 gas supply. When cells other than single sperm were analysed, needle hole diameter was increased to between 10 and 20 μm. Cells were transferred to a drop of mineral oil in the cap of 0.5 ml Eppendorf tubes for DNA extraction and subsequent PCR amplification. Two and a half microlitres of 200 mM KOH and 50 mM DTT was added to the cells and they were incubated for between 10 and 30 min at 65°C. This solution was neutralised with 2.5 μl neutralising buffer (300 mM KCl, 900 mM Tris (pH8.3), 200 mM HCl) (Cui *et al.*, 1989).

The 5 μl solution containing the lysed cell was directly subjected to a PCR amplification of one cycle of 94°C/3 min, 40 cycles of 94°C/15 sec, 50°C/15 sec, and 72°C/1 min, and one cycle of 72°C/5 min using primers HMTL822 (5'ATTCCCCTAAAAATCTTTGAAA) and HMTH897 (5'tgagtaggctgatg). This 25 μl reaction contained 2.5 pmol of each primer, 2.5 nmol of each dNTP, 2.5 μl of 10 x reaction buffer (500 mM KCl, 100 mM Tris-HCl (pH 9.0)), 2 mM $MgCl_2$ and 0.5 units of *Taq* polymerase (Boehringer Mannheim). A 5 μl aliquot of this primary 595 bp PCR product

was then re-amplified using a hemi-nested PCR protocol. Primer HMTH897 was joined by the internal primer HMTL854 (5'TCATTCATTGCCCC) to amplify a 430 bp fragment using the same cycling conditions as above. Primer quantities were increased to 5 pmol and the $MgCl_2$ concentration for this internal PCR was 1 mM. Conditions for the above PCR reactions are based on Lien *et al.* (1993). A 1.5 µl aliquot of the second PCR product was directly analysed by SSCP analysis as before. In addition to single cells and cellular aggregates being analysed in this way, seminal fluid controls were taken. Five microlitres of seminal fluid suspended in PBS was centrifuged for 10 min at 12,000 x g to pellet all cells. That there were no cells in suspension was checked in a Makler counter (Makler, Israel). Aliquots of seminal fluid were then treated as before to prepare DNA and subjected to the same PCR reactions and SSCP analysis.

Cycle sequencing of the 430 bp PCR product and other PCR products containing homoplasmic changes was carried out using a Gibco BRL dsDNA cycle sequencing system with primers HMTL817, HMTL854, HMTH897 and HMTH934. To obtain purified PCR product for sequencing, products from a number of PCR reactions were pooled. An equal volume of 3 M ammonium acetate was added, followed by two volumes of isopropanol. DNA was precipitated for 10 min at room temperature before being pelleted by centrifugation at 12,000 x g for 10 min at room temperature. The DNA pellet was washed with 500 µl 70% ethanol, spun as before, air dried and resuspended in TE8. One picomole of the appropriate primer was isotopically end labelled with $\gamma^{32}P$ or $\gamma^{33}P$ ATP (~3,000 Ci/mmol) and included in the appropriate reaction with 50 fmol of template DNA. The PCR conditions used for sequencing were 20 cycles of 95°C/30 sec, 50°C/30 sec and 70°C/1 min followed by 10 cycles 95°C/30 sec and 70°C/1 min. Sequencing gels were pre-run at 1500 V, 35 - 45 mA for 1 hour. Individual sequencing mixes were heated to 80°C for 2 min and rapidly cooled on ice. Two microlitres of denatured sample was used for sequencing. Gels were run at 1500 - 1700 V with a constant current of 35 mA per gel in a BRL S2 sequencer. After electrophoresis the gels were dried onto backing paper in a Biorad 583 gel drier and exposed to X-ray film (Kodak X-OMAT AR) for different lengths of time depending on their radioactive counts. Films were developed and the sequence was scored manually.

Results

Six different SSCP patterns were seen from the screening of the *Taq1* cleaved ATPase amplicon from 210 semen samples (Table 2.1). The most common pattern (not shown in Table 2.1) is assumed to represent the Cambridge sequence as this is a consensus sequence of European origin. For a full list of screened samples see Appendix 1.

Of the 210 samples screened, 31% were oligozoospermic 48% of these being severely oligozoospermic ($< 5 \times 10^6$ sperm/ml) (excluding missing data), 49% of samples were asthenozoospermic. A strong correlation exists between reduced sperm count and motility in this sample group (S. Richards, unpublished Masters Thesis, University of Canterbury, 1996). A number of individuals have known seminal defects including Y chromosome micro-deletions (N. Kerr, unpublished Masters Thesis, University of Canterbury, 1997). Additionally, not all seminal samples contain the same cell types. A number of samples were from individuals who have undergone vasectomies reducing the cells in the semen to peripheral leucocytes and possible epididymal cells. There were a number of samples that were obtained from seminal swimup or gradient procedures. These procedures enrich the semen sample in motile sperm.

Pattern 1 (Fig. 2.3a) within sample EV, a drop in the bottom 789 nt band, is indicative of a homoplasmic change. This pattern was reproducible from different PCRs from the same semen sample but was not found in any other semen sample. When the EV ATPase amplicon was cleaved with *Apa*1 the bands representing the 1016 bp fragment showed a mobility shift when compared with control semen samples (Fig. 2.3b). Bands representing fragments 753 bp and 558 bp, generated by cleavage with *Sca*1 and *Hin*f1 respectively, also showed mobility shifts when compared with control samples (Fig. 2.3b). The overlapping region between all of the fragments showing mobility shifts was between the *Hin*f1 cleavage site at nt 8784 and the *Taq*1 cleavage site at nt 8958. Therefore, it was hypothesised that pattern 1 was caused by a homoplasmic change between nt 8784 and nt 8958.

Table 2.1. SSCP patterns that differed from the control pattern after *Taq*1 digestion of the ATPase amplicon of 210 semen samples.

Pattern	Sample	Fragment	Change
1	EV	789	bottom 789 band dropped
2	QD747V	789	both 789 bands move
3	SF95-171	789	top 789 band drops
4	94-149	384	bottom 384 band raised
4	95-5	384	bottom 384 band raised
4	95-17	384	bottom 384 band raised
4	95-22	384	bottom 384 band raised
4	95-34	384	bottom 384 band raised
4	GB1	384	bottom 384 band raised
4	MC	384	bottom 384 band raised
4	QD720I	384	bottom 384 band raised
4	QD8070	384	bottom 384 band raised
4	QR511C	384	bottom 384 band raised
4	SF362	384	bottom 384 band raised
4	SF94-86	384	bottom 384 band raised
4	TI	384	bottom 384 band raised
5	QR568B	789	extra heteroplasmic band

Full length EV ATPase amplicon was prepared in duplicate and the region of interest was sequenced using primers HMTL854 and HMTH897. A G → A transition at nt 8860 was the only change seen in both amplification and sequencing reactions and was seen in all sequenced strands (Fig. 2.3c). This transition changes a threonine codon (ACA) to an alanine codon (GCA) therefore changing the amino acid assignment at residue 112 of the ATPase 6 polypeptide. Given that the G → A transition at nt 8860 was the only change seen it was concluded that this transition created pattern 1.

Patterns 2 (Fig. 2.4a) and 3 (Fig. 2.5a) within semen samples QD747V and SF95-171 respectively, had mobility changes in the 789 nt *Taq*1 bands when compared with the control pattern. Patterns 2 and 3 were reproducible from PCRs from the DNA extractions from the same semen samples but were not found in any other samples. Cleavage of the QD747V and SF95-171 ATPase amplicons with *Hin*f1 and *Rsa*1 and subsequent SSCP of the resultant fragments showed changes in band position

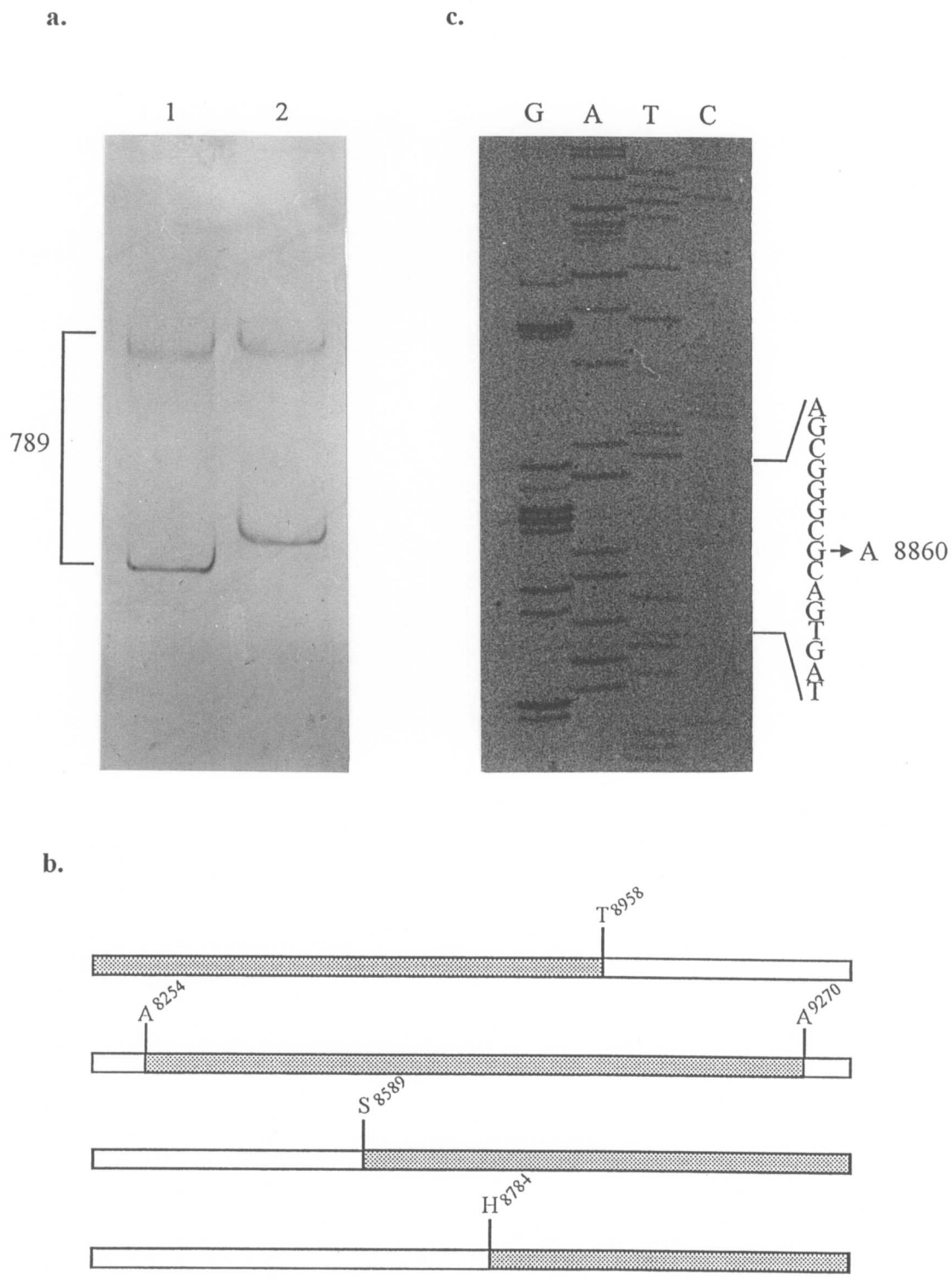


Figure 2.3 **a**. Pattern 1 from ATPase amplicon screening of semen sample EV. Lane 1 shows the two bands representing the 789 bp fragment from *Taq*1 cleaved EV ATPase amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). **b**. Schematic representation of restriction endonuclease cleavage sites within the ATPase amplicon for enzymes used in the analysis of pattern 1. Restriction sites: *Taq*1 (T), *Apa*1 (A), *Sca*1 (S), *Hin*f1 (H). Fragments showing mobility difference with respect to control samples are shaded. **c**. Portion of cycle sequencing gel showing the G → A transition at nt 8860. This sequence was derived from primer HMTH987. Compare this sequence with Fig. 2.5**b**.

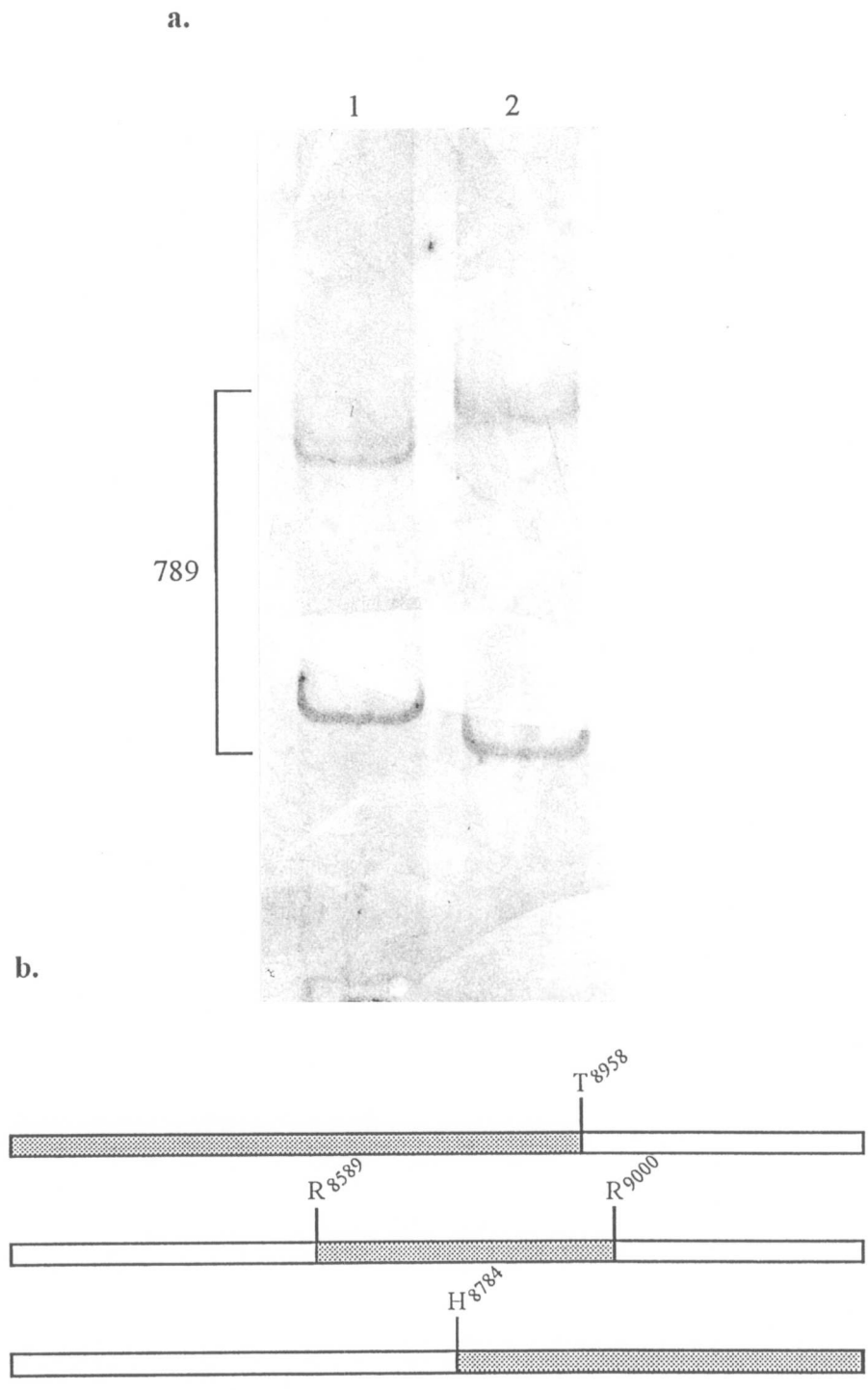


Figure 2.4 **a**. Pattern 2 from ATPase amplicon screening of semen sample QD747V. Lane 1 shows the two bands representing the 789 bp fragment from *Taq*1 cleaved QD747V ATPase amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). **b**. Schematic representation of restriction endonuclease cleavage sites within the ATPase amplicon for enzymes used in the analysis of patterns 2 and 3. Restriction sites: *Taq*1 (T), *Hin*f1 (H), *Rsa*1 (R). Fragments showing mobility difference with respect to control samples are shaded.

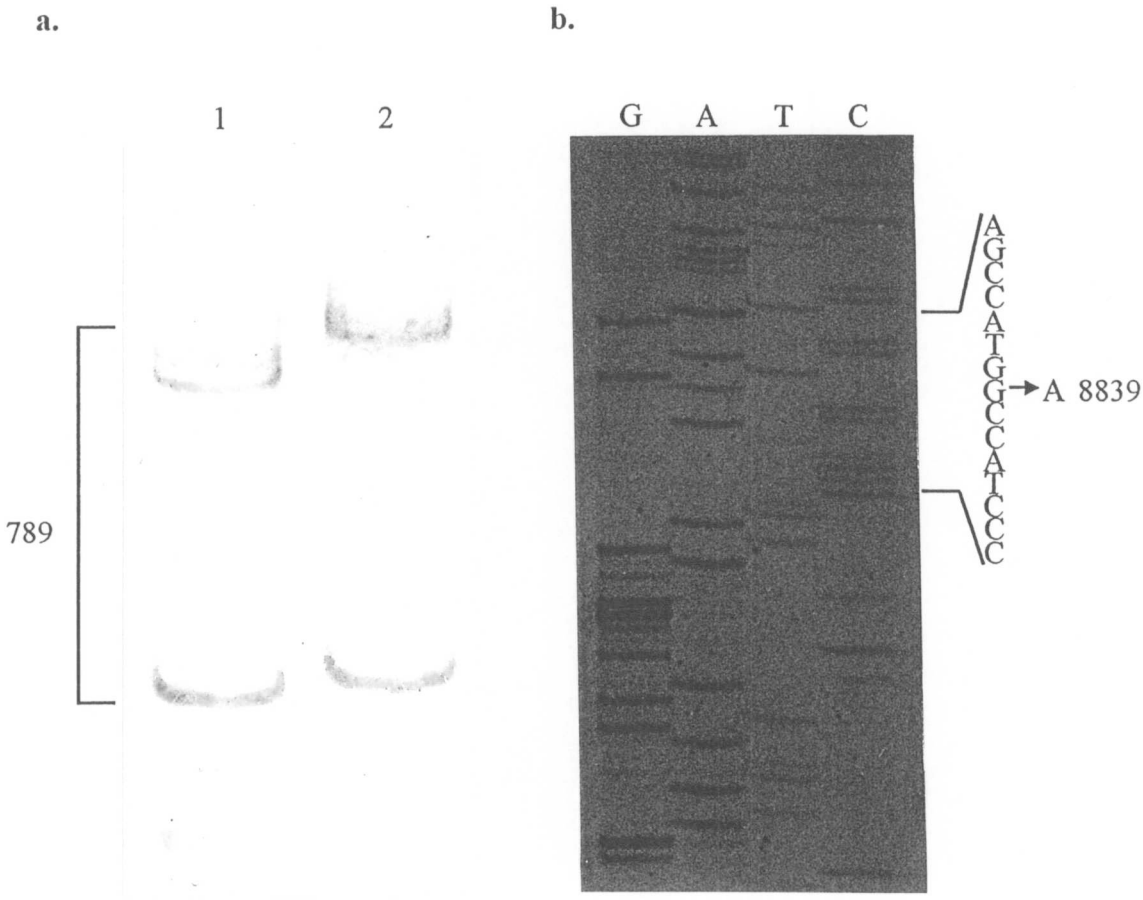


Figure 2.5 **a**. Pattern 3 from ATPase amplicon screening of semen sample SF95-171. Lane 1 shows the two bands representing the 789 bp fragment from *Taq*1 cleaved SF95-171 ATPase amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). **b**. Portion of cycle sequencing gel showing the G → A transition at nt 8839. This sequence was derived from primer HMTH987. Compare this sequence with Fig. 2.3**c**.

corresponding to fragments of 558 bp and 411 bp respectively (Fig. 2.4b). Therefore, the overlapping region between all of the fragments showing mobility shifts is between the *Hin*f1 cleavage site at nt 8784 and the *Taq*1 cleavage site at nt 8958 for both patterns 2 and 3.

Full length QD747V and SF95-171 ATPase amplicons were prepared in duplicate PCRs and the region of interest was sequenced using primers HMTL854 and HMTH897. A G → A transition at nt 8856 was the only change seen in duplicated amplification and sequencing reactions from QD788L and was seen in all sequenced strands (result not shown). This mutation changes an alanine codon (GCG) to another alanine codon (GCA) therefore not changing the amino acid assignment at residue 110 of the ATPase 6 polypeptide. Given that the G → A transition at nt 8856 was the only change seen it was concluded that this created pattern 2. A G → A transition at nt 8839 was the only change seen in both amplification and sequencing reactions of SF95-171 and was seen in all sequenced strands (Fig. 2.5b). This mutation changes an alanine codon (GCC) to a threonine codon (ACC) thereby changing the amino acid assignment at residue 105 of the ATPase 6 polypeptide. Given that the G → A transition at 8839 was the only change seen it was concluded that this created pattern 3. This change also created a *Hae*III RFLP with the loss of the *Hae*III restriction site centred at nt 8840.

Pattern 4 (Fig. 2.6), a change in the mobility of bands representing the 384 bp *Taq*1 fragment, was seen in 13 semen samples. This pattern was indicative of a homoplasmic change. When the screening of semen samples showing pattern 4 was repeated, this pattern was frequently not seen again indicating that the mobility shift with respect to controls was very condition specific. Because of the variability in band shifting it is likely that 13 semen samples with this pattern is an underestimate. When semen samples with pattern 4 and controls were cleaved with *Hae*II, those with pattern 4 had a *Hae*II RFLP due to the loss of a restriction site at position nt 9056 (not shown). The loss of this site has been seen previously and defines European haplotype K. European haplotype K is found in 7.4% of Caucasians (Torroni *et al.*, 1994) and has been seen in all cited Caucasian populations studied to date. No study has defined the common nucleotide change that leads to this haplotype, however. This study is no exception. Marzuki *et al.* (1991) and Noer *et al.* (1991) both describe a G → A

transition at nt 9055 which would lead to this RFLP but neither make the connection to European haplotype K. No other polymorphisms have been characterised in this *Hae*II restriction site (RGCGCY). The change causing pattern 4 in the present study therefore could be a G → A transition at nt 9055 but this has not been confirmed. Seven of the 13 samples with pattern 4 were oligozoospermic. This was not statistically significantly different from the proportion of oligozoospermic samples in the sampled population (exact binomial probabilities, p = 0.0488, two tailed test with confidence limits of 95%). Additionally 9 of the 13 samples were asthenozoospermic. Again this was not significantly different from the proportion of asthenozoospermic samples in the sampled population (exact binomial probabilities, p = 0.0789, two tailed test with confidence limits of 95%).

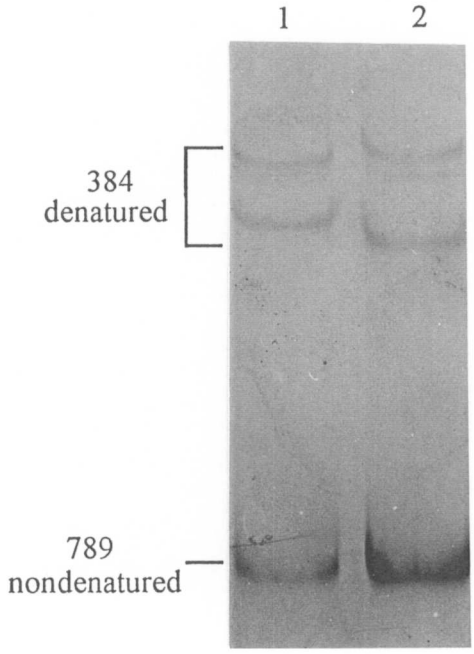


Figure 2.6. Pattern 4 from ATPase amplicon screening. Lane 1 shows the two bands representing the 384 bp fragment from *Taq*1 cleaved QR511C ATPase amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence).

Pattern 5 (Fig. 2.7a) is found only in semen sample QR568B and a repeat semen sample (QD205F) from the same individual. This pattern is characterised by an extra band between the *Taq*1 generated 789 bp fragment bands. The presence of an extra band in addition to normal bands is indicative of heteroplasmy. Samples QR568B and QD205F were donated 9 months apart. The donor, at the time of donation was a 26 year old healthy

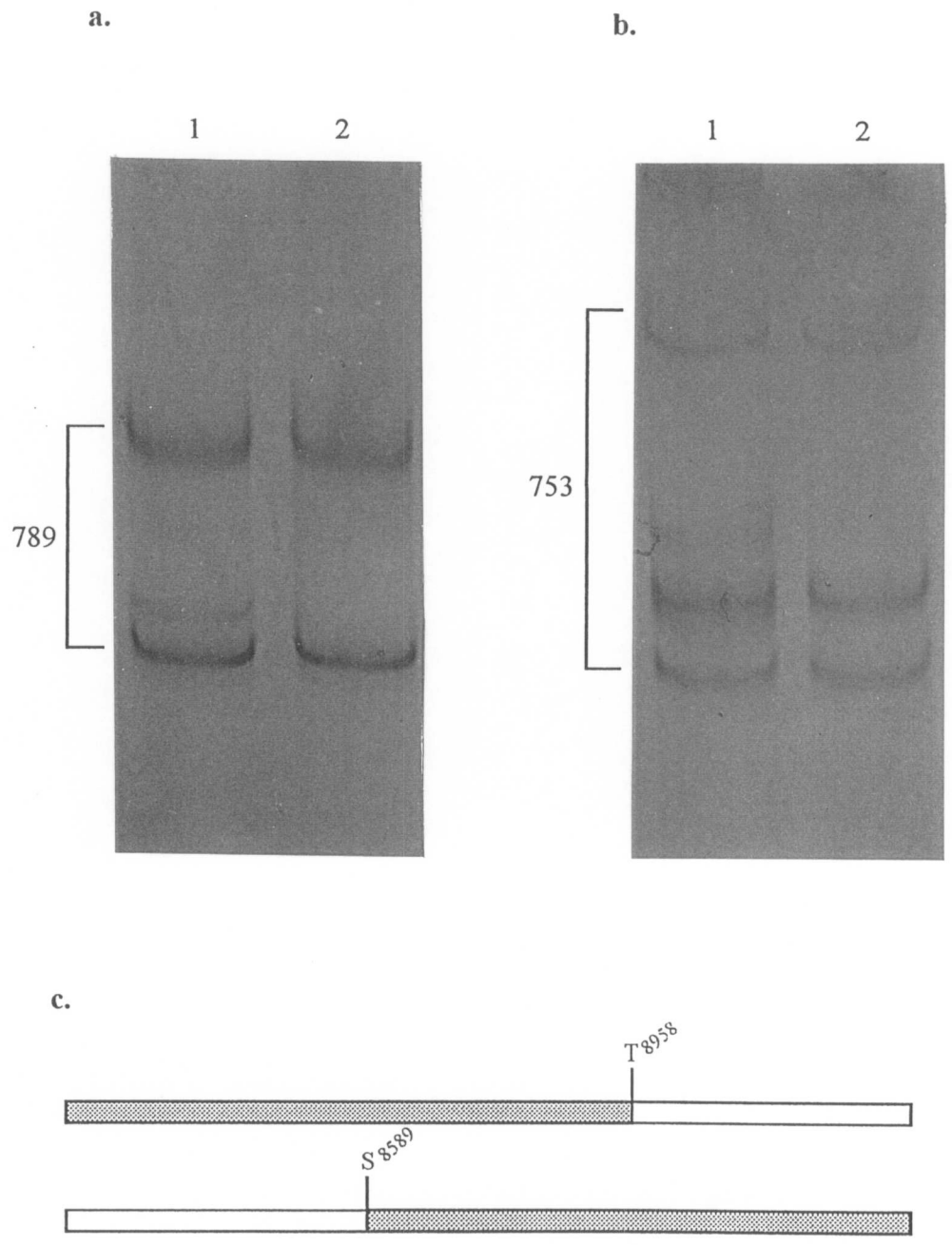


Figure 2.7 **a**. SSCP analysis of QR568B (lane 1) and a control individual (lane 2) showing the presence of the heteroplasmic band within a 789 bp *Taq*1 digest fragment. **b**. SSCP analysis of QR568B (lane 1) and a control individual (lane 2) showing the presence of the heteroplasmic band within a 753 bp *Sca*1 digest fragment. Note that the 753 bp *Sca*1 fragment runs as three bands, instead of the usual two. **c**. Schematic representation of restriction endonuclease cleavage sites within the ATPase amplicon for enzymes used in the analysis of pattern 5. Restriction sites: *Taq*1 (T), *Sca*1 (S). Fragments showing mobility difference with respect to control samples are shaded.

male with primary infertility due to oligo-astheno-teratozoospermia. He had a normal blood follicle stimulating hormone level and the two semen samples analysed were: 4.3 ml with a count of 0.3 x 10^6 sperm/ml, 20% motility; and 4.6 ml with 0.9 x 10^6 sperm/ml, and 17% motility. Both samples had large numbers of immature forms (spermatids), and a reduced forward progression (I.L. Sin, pers. comm). Only the second sample (QD205F) had enough sperm for a morphology assessment, and it showed 100% abnormal heads, as assessed by WHO criteria (1992). Both samples tested negative for a peroxidase test indicating that there are fewer than 1 x 10^6 peripheral blood lymphocytes per ml of semen. Scanning densitometry of SSCP gel bands from 12 different PCR reactions from either of the two DNA samples indicates that 16% of the PCR amplified DNA contains the mutation (x = 16.3%, s = 2.5%). An analysis of the peripheral lymphocytes from this donor also revealed the presence of the same heteroplasmic mutation, with 18% mutant DNA (x = 17.8%, s = 6.36%). Silver staining is, however, only semi-quantitative so these estimates have errors of unknown magnitude associated with them.

The mutation causing pattern 5 was narrowed down by cleavage of the PCR product with *Sca*1 and subsequent SSCP. The heteroplasmic mutation was in a 753 bp *Sca*1 fragment (Fig. 2.7b). Double digestion of the QR568B ATPase amplicon with *Taq*1 and *Sca*1 delineated the heteroplasmic mutation to a 369 bp fragment between nt 8589 and nt 8958 with no heteroplasmy being seen in any of the other fragments (Fig. 2.7c).

Oligonucleotide primers HMTL854 and HMTH897 were designed, flanking this region, with the resultant 430 bp PCR product containing the heteroplasmy when tested by SSCP analysis. In an effort to obtain an enriched sample of mutant DNA for sequencing, single cell analysis was initiated. Semen samples QR568B and QD205F contained a range of cells other than mature sperm. These included epithelial cells, red blood cells, white blood cells, and various forms of immature sperm (spermatogonia, spermatocytes and spermatids). Identification of these cells is very difficult but is often clinically important. Neutrophils (white blood cells) can be determined, when live, by the peroxidase staining test but this method is crude and inaccurate. Other than this, electron microscopy is the only reliable method of positively identifying these cell types (Glover *et al.*, 1990). Additionally, in semen samples there are often high titres of antisperm antibodies. These can instigate clumping (head to head or tail to tail) of sperm and the

agglutination of sperm to other cell types as is seen in QR568B and QD205F. Because of the confusion over cell identity, cells were categorised into classes based on size. The classes and the most probable cell types that they represent are listed in Table 2.2. Cell counts of QD205F indicate that there are high numbers of large round cells (9.4×10^6 /ml (n = 4)) and even higher numbers of small round cells (40.2×10^6 /ml (n = 4)).

Two hundred and fifty four single cells and aggregates were analysed by PCR. Ninety five of these gave positive PCR signals and were further analysed by SSCP. The results of these are summarised in Table 2.3 with the raw data appearing in Appendix 3. Two from 30 seminal fluid controls gave positive PCR results, both having a normal SSCP pattern. This indicated that there was some extra cellular DNA in solution.

Table 2.2. Classed cell types found in semen samples QR568B and QD205F. Classes were used in single cell analysis, when individual cells are unidentifiable.

Class	Description	Probable cell types
I	tiny, spherical transparent bodies	possibly cytoplasmic droplets?
II	single small round cells	spermatids
III	single large round cells	Epithelial cells, spermatogonia, spermatocytes and leucocytes
IV	single sperm	single sperm
V	single large irregular cells	macrophages (?)
VI	clumped cells other than sperm	clumped class II and/or III
VII	clumped sperm	clumped sperm often with spermatids

Table 2.3. Summarised single cell and aggregate data for cells that gave positive PCR signals. *; mutant patterns with negligible amounts of normal DNA that were used for sequencing.

Class	Number analysed	SSCP pattern		
		normal	mutant	mixture
I	-	-	-	-
II	16	15	1	0
III	14	14	0	0
IV	32	31	0	1
V	9	8	0	1
VI	11	9	0	2
VII	13	10	2*	1

As shown in Table 2.3 there are two examples of nearly completely mutant SSCP patterns from class VII aggregates (Fig. 2.8a). The PCR samples giving rise to these were sequenced by PCR cycle sequencing using primers HMTH897 and HMTL854. Both of these sequences identified a T → C transition at nt 8821 (Fig. 2.8b, c). This was the only change found in both PCR reactions and these PCR products showed mutant SSCP patterns. It was therefore hypothesised that the heteroplasmic pattern seen in samples QR568B and QD205F was caused by a T → C transition at nt 8821. The class II cell with a wholly mutant pattern (Table 2.3) was sequenced and found to have a different change. Given that this was seen only once, and the SSCP pattern created was neither normal nor the traditional mutant (Appendix 3) it is assumed that this change was caused by PCR artefact. Sequence from the PCR product of three single sperm was the same as the Cambridge sequence. A T → C transition mutation is predicted to change a serine codon (TCT) to a proline codon (CCT) and hence change Ser^{99} to a proline in the ATPase 6 polypeptide.

Discussion

Sensitivity of PCR-SSCP for mutation detection

Three different homoplasmic SSCP patterns (1 - 3) were caused by changes in the 789 bp *Taq*1 bands. These changes all fall within 40 bp of each other. Rather than this being a hotspot for mutations in mtDNA it is likely that this represents extremely low sensitivity in the mutation detection system. The main problems are probably the extreme size of the 789 bp *Taq*1 fragment and the lack of different gel electrophoresis conditions. The effect of size of the DNA fragment on SSCP sensitivity is much debated. The optimum fragment size is predicted to be between 100 and 300 bp, with fragments of 400 - 500 bp having a detection sensitivity of between 70 and 80% (Beier, 1993). Fragments of 600 bp have been analysed by SSCP with a sensitivity of 3% (Sheffield *et al.*, 1993) with the authors conceding, however, that the sequence context plays a very large part in this (different sequences of 135 bp having sensitivities of between 90 and 59%). Additionally Sheffield *et al.* (1993) attempted to detect variation in large fragments using SSCP conditions optimised for small fragments. There are a number of ways around the problem of lack of sensitivity with increased fragment size. Firstly, the ATPase amplicon could have been divided up into smaller overlapping amplicons (200 - 300 bp in length) that could have been

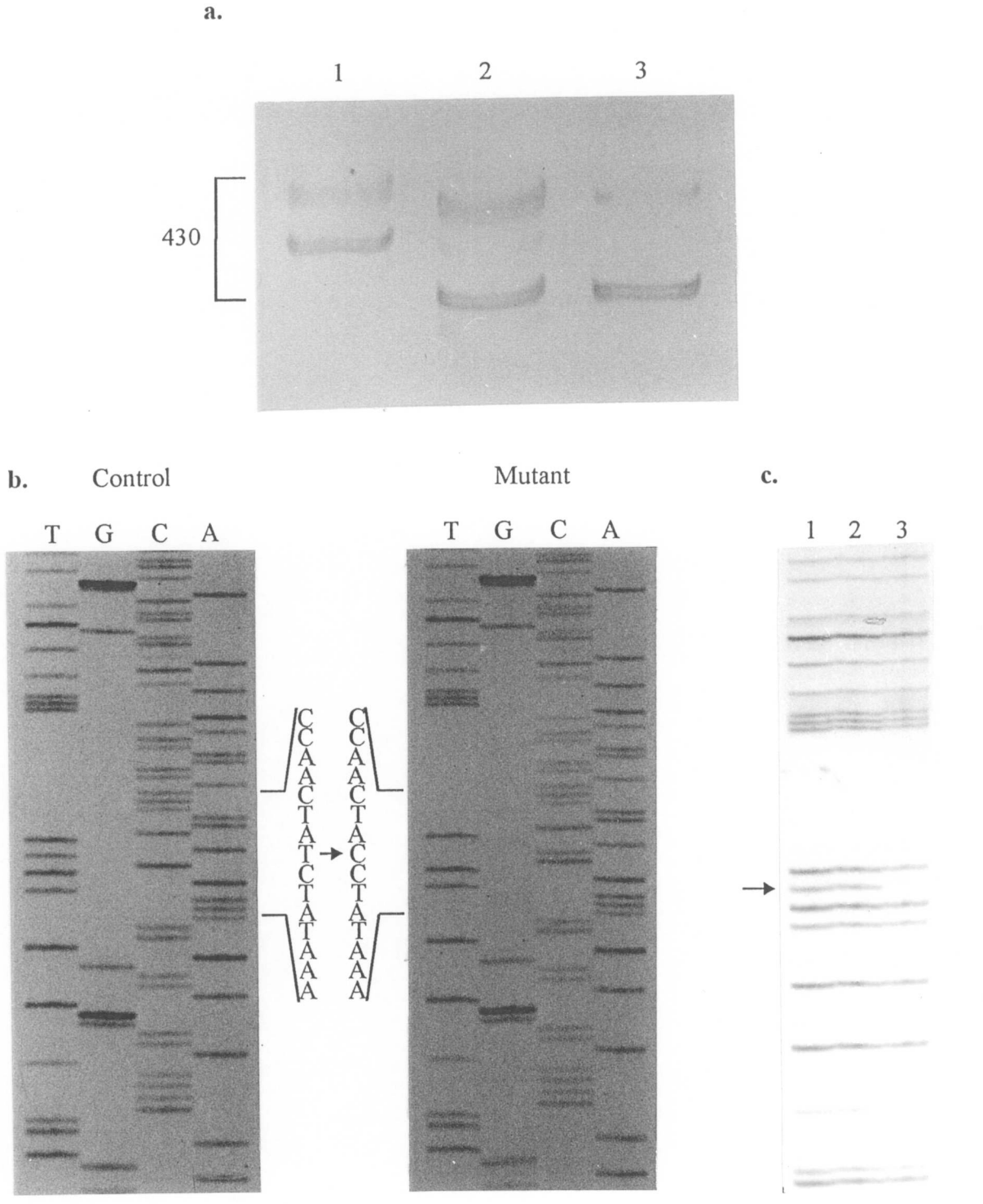


Figure 2.8 **a**. SSCP analysis of 430 bp PCR fragment derived from oligonucleotides HMTL854 and HMTH897. Lane 1, homoplasmic mutant DNA amplified from immature spermatids from QR568B. Lane 2, heteroplasmic SSCP pattern from QR568B. Lane 3, homoplasmic normal DNA amplified from single mature sperm from QR568B. **b**. Double stranded PCR sequence analysis using a 430 bp PCR fragment of DNA amplified from single mature sperm (control) and cells within a clump of immature spermatids (mutant) from QR568B. The control sequence is derived from the PCR product shown in Fig. 2.8 **a** (lane 3) whilst the mutant sequence is derived from the PCR product shown in Fig. 2.8 **a** (lane 1). The T → C transition at nt 8821 is indicated. **c**. T-tracking illustrating the absence of the mutation in a control individual (lane 1), and from single mature sperm from the QR568B (lane 2). Lane 3 is the sequence from the immature spermatids. The site of the mutation is marked by the arrow. All control sequences were the same as the Cambridge sequence (Anderson *et al.*, 1981).

analysed individually. Because these amplicons are required to be overlapping (for full gene coverage) adjacent amplicons could not be multiplexed. The requirement, therefore, for more than one amplification reaction does not make this a cost effective option. Additionally with more primers, amplifying short sequences, there is more chance of allele specific amplification because of a primer binding over a mutation site. Secondly, longer PCR products could be generated and cleaved with different enzymes to create smaller fragments for analysis. Whilst this is good in theory it is not practical because large tracts of the ATPase amplicon have no restriction enzyme recognition sites. Additionally suitable restriction enzymes are often prohibitively expensive for routine screening. In the present situation *Rsa*1 cleaves the ATPase amplicon at two positions giving fragments of 342 bp, 411 bp and 421 bp. Unfortunately the SSCP bands of these fragments frequently could not be separated from one another because of their sizes.

The sensitivity of SSCP is dependant on conditions such as the size of the DNA fragment to be analysed, the gel temperature, ionic strength of the electrophoresis buffer and the porosity and composition of the gel (Hayashi, 1992 and references therein). Whilst the detection of changes in different DNA fragments often depends on these factors it is impractical to screen all samples with a variety of different conditions. Therefore one set of optimised conditions is used, with the possibility of some loss of sensitivity. Ideally these conditions are tested with a set of known mutations, but where these are not available an empirical approach is adopted.

Other mutation detection systems can and have been used to analyse mtDNA mutations. Broadly speaking there are two groups of methods used for mutation detection. Diagnostic methods are used to detect known mutations whilst scanning methods are designed to search for new mutations. For these methods direct sequencing is the most informative but is expensive. Thus sequencing is used only as the last step in mutation identification. Most diagnostic methods use the principle of specific oligonucleotide binding, whether as isotopically labelled probes or allele specific primers for PCR amplification (eg. Gibbs *et al.*, 1989; Seibel *et al.*, 1994). Additionally RFLP analysis is still commonly used where the mutation falls on a known restriction endonuclease recognition site.

Scanning methods have been used successfully to detect changes in mtDNA where all have their inherent strengths and weaknesses. Most of the known polymorphisms in human mtDNA have first been seen as RFLP changes. However, using a high intensity RFLP procedure of 12 restriction enzymes, only approximately 20% of mtDNA will be screened, although all large insertions and deletions will be found (Brown and Wallace, 1994). Whilst this technique has the ability to screen thousands of bases of DNA and accurately pin point changes, the lack of recognition sites and excessive cost remain prohibitive. Many of the other methods rely on the separation of hetero- or homoduplexes in polyacrylamide gels.

Denaturing gradient gel electrophoresis (DGGE) relies on the differential separation of partially melted homo- and heteroduplexes when electrophoresed through a gradient of denaturants (such as formamide or urea) (Fischer and Lerman, 1983). Frequently this method requires a high T_m G-C clamp to be attached to one end of the sequence to prevent complete strand separation and allow detection of base changes throughout the whole sequence (and not just in the lowest T_m regions) (Myers *et al.*, 1985). Additionally the need for high temperatures and denaturant gradients makes this procedure both costly and time consuming (Lessa and Applebaum, 1993). Variations on this technique include temperature gradient gel electrophoresis (Wartell *et al.*, 1990) and constant denaturant gel electrophoresis (Smith-sørenson *et al.*, 1992) both of which have similar limitations to DGGE. Methods based around DGGE can have sensitivities approaching 100% for DNA fragments up to 500 bp in length (Lessa and Applebaum, 1993).

The Heteroduplex method (HET) simply relies on the differential mobilities of hetero and homoduplexes in polyacrylamide gels. HET has the ability to show heteroplasmy directly and, with doping of samples, to detect homoplasmy. Whilst no special equipment is required, and this method is very simple to perform, the sensitivity is often reduced and the ability to differentiate between two different changes is frequently absent (Lessa and Applebaum, 1993).

Enzymatic methods utilising the ability of S1 nuclease and RNase A to cleave single stranded DNA and RNA respectively have been developed. S1 nuclease can theoretically cleave single stranded DNA loops at sites of mis-matches within heteroduplexes but with single base mis-matches this ability is almost negligible (Silber and Loeb, 1981). RNAse A

can cleave at single base mis-matches with high affinity and has 60 - 70% sensitivity but is expensive and requires the production of an RNA probe (Dianzani *et al.*, 1993). Modifications of this method include the chemical cleavage method (CCM) in which mis-matched T and C bases can be reacted with osmium tetroxide and hydroxylamine respectively before being cleaved with piperidine (Cotton, 1989). The fact that the sensitivity of this method is unknown and toxic chemicals are involved does not make this a method of first choice.

Mutation analysis

Six different SSCP patterns were seen from the screening of 210 semen samples for changes in the ATPase amplicon. The most common of these is assumed to represent the Cambridge sequence. Only pattern 4, found in at least 13 semen samples, was tested for statistically significant linkage to seminal parameters. Seven of the 13 samples with this change were oligozoospermic while 9 were asthenozoospermic. Neither the sperm count nor percent sperm motile from pattern 4 semen samples were significantly different from the sampled population proportions (exact binomial probabilities, $p > 0.025$, two tailed test with confidence limits of 95%).

The G → A transition at nt 8860 within semen sample EV has been characterised previously. The original Cambridge sequence (Anderson *et al.*, 1981) has an A at this position with subsequent studies showing that a G in this position is probably more common among Europeans. Thus nt 8860 is considered to be one of the "mistakes" in the Cambridge sequence (see Chapter 1). A number of other studies have noted that whilst nt 8860 is probably a G, there are individuals who have an A in this position (e.g. Shoffner *et al.*, 1993, individuals C8 and C12). Frequently this is noted by the loss of an *Hha*1 restriction site. The population frequency of an A at nt 8860 is not known. In the present study the population frequency of this change was 0.48%. The nucleotide change at this position changes a threonine codon (ACA) to an alanine codon (GCA) therefore changing the amino acid assignment at residue 112 of the ATPase 6 polypeptide. No function has been assigned to this amino acid position and its homology is not well conserved (threonine in human and *Xenopus*, alanine in Bovine and mouse (Roe *et al.*, 1985)).

The G → A transition at nt 8856, in this study seen in semen sample QD747V, has been characterised previously (Ozawa *et al.*, 1991; Tanaka and Ozawa, 1994). Tanaka and Ozawa (1994) found this change in two out of 43 Japanese and Australian individuals. In the present study the population frequency of this change was 0.48%. Given that this change does not alter amino acid assignments it is assumed to be a neutral or nearly neutral population variant. No information is known about QD747V's donor ethnicity so it is not possible to say if this variant has occurred just once or more frequently within recent human history.

The G → A at nt 8839, found within semen sample SF95-171, has not previously been characterised. A *Hae*III RFLP centred around this site has been seen before (Wallace *et al.*, 1995) but the nucleotide causing this RFLP has never been delineated. It is probable that nucleotide changes in addition to the G → A transition at nt 8839 can cause this RFLP. Interestingly, Ala^{105} is reasonably well conserved having been seen in all vertebrates to date and *Drosophila* but not in honey bees or sea urchins. Given that a *Hae*III RFLP covers this site, and this RFLP has been seen in humans before, it is unlikely that this change whilst altering a relatively highly conserved nucleotide, has any pathological significance. Being found in only one out of 210 semen samples the population frequency of this change was 0.48%.

The uncharacterised *Hae*II RFLP seen in 13 samples is a marker for European haplotype K. Torroni *et al.* (1994), who first specified this change as a marker based on phylogenetic analysis, claim that it is present in 7.4% of Caucasians. In certain populations such as the Ashkenazi Jews this frequency can be up to 30% (Ritte *et al.*, 1993). The estimate of 6.2% (13/210) in this study is probably an under representation. Slight variability in SSCP gel conditions may have contributed to the mobility shift in some gels but not in others. The value of 6.2% in this study is not significantly different from the population percentage (7.4%) calculated by Torroni *et al.* (1994) ($\chi^2 = 0.410$, df = 1, $P > 0.05$).

The heteroplasmic T → C transition at nt 8821 has not been characterised previously. No restriction endonuclease recognition site encompasses this position in either its normal or mutated state. This change was shown to be heteroplasmic in two semen samples from one

individual donated 9 months apart. No other semen samples have this change. At the time of the second donation a blood sample also contained the heteroplasmy with approximately the same proportion of mutant DNA as semen (see results). No familial studies were possible, within the bounds of this study, to analyse the familial nature of this mutation. It is unknown, therefore, whether the mutation in this individual has been inherited by him or arose within his mothers egg or his early blastocyst. The detection of this mutation within small round cells (most probably immature spermatids) negates the suggestion that its source in the semen is from lymphocyte contamination.

Determining the pathogenicity of a mtDNA mutation is difficult. One of the key criteria, as stated by Riordan-Eva and Harding, (1995) is that heteroplasmy be present, as heteroplasmy is rarely a feature of harmless polymorphisms. The proportion of mutant DNA in a tissue has been shown to correlate with the severity of the phenotype associated with many heteroplasmic mtDNA mutations. For example a high level (>90%) of a T $\rightarrow$ C or T $\rightarrow$ G at nt 8993 within the ATPase 6 gene causes Leigh's disease, whilst a lower level (between 70% and 80%) may cause only peripheral retinopathy (Brown and Wallace, 1994). Where heteroplasmy is involved it is believed that seminal fluid can have similar properties to somatic tissues. However, the pathogenicity of mtDNA mutations is dependant on the threshold, and the cell function. The more mature, active, healthy spermatozoa that are ejaculated, the greater the chance there is of a fertilisation event. If a heteroplasmic mutation were to affect the motility of a large proportion of spermatozoa or decrease the number of mature sperm in ejaculated seminal fluid then it could be considered to be pathogenic in the same way as that occurs in somatic tissues, where tissues have a threshold of mutations they can withstand before exhibiting phenotypic effects. Important here is the concept of inter- versus intra-cellular heteroplasmy. Tissues may have a higher mutational threshold if mutations are spread over all cells, with each cell having both normal and mutant DNA (intra-cellular heteroplasmy) than if individual cells are either normal or mutant (inter-cellular heteroplasmy). Examples of both types of heteroplasmy can be seen in the literature. Studies of cloned myoblasts from patients carrying a MERRF mutation (mtDNA $tRNA^{Lys}$ A $\rightarrow$ G at nt 8344) show that the clones are homoplasmic for either wild-type or mutant mtDNA (Boulet *et al.*, 1992; Larsson *et al.*, 1992). Contrary to this is the finding that patients with MELAS can have intra-cellular heteroplasmy in fibroblasts which is maintained with cloning (Matthews *et al.*, 1995). Whilst it was originally thought that there was a progression from

intra- to inter-cellular heteroplasmy with repeated random assortment of mtDNA, there are examples of heteroplasmy maintenance, far beyond what would be expected from random assortment. Where the mutation is deleterious it is advantageous to retain heteroplasmy because homoplasmic mutants are often lethal. This then begs the question of whether sperm are individually homo or heteroplasmic. To appreciate the consequences of this, consider a semen sample with a 50% mutational load. If this 50% is placed in all cells evenly, then the sperm will probably survive, perhaps with a reduced forward progression but still with normal motility. If the 50% mutant load is in 50% of the cells then the effect on fertilising potential could be quite marked as there is an effective drop in sperm motility of 50%. This study does not really address this issue for a number of reasons. Firstly, all examples of mutant DNA found in single cell or aggregate analysis had at least traces of normal mtDNA. It is tempting to say that this is due to extra cellular DNA contamination but intra-cellular sources can not be ruled out. Secondly, even if cells appeared to be individually homoplasmic, they may be so because of allele drop-out (ADO). ADO occurs when the amount of template DNA is very small and one allele is amplified preferentially to another leading to a false homozygous (in this case homoplasmic) state (Ray *et al.*, 1995). By using aggregates of cells, the chance of ADO is reduced.

The stability of the mutation in two semen samples donated 9 months apart suggests that the precursor germ cells may also carry the same mitochondrial mutation. One scenario is that the small number of fully mature sperm, possibly without the mutation, may be derived from unaffected germ cells, and the immature spermatozoa containing the mutation may be derived from the precursor cells carrying the mutation. A second scenario is that the precursor germ cells are individually heteroplasmic. If this is true, then it would suggest non-random segregation of the mtDNA during spermatogenesis, that is, the segregating units of mtDNA during meiosis are homoplasmic. A number of studies are presently looking at the clonal nature of mtDNA in sperm (A. Jaquier, pers. comm.; J. Poulton, pers. comm.). At present it is not possible to differentiate between the two scenarios.

The T $\rightarrow$ C transition mutation is predicted to cause the replacement of Ser^{99} with proline in the ATPase 6 polypeptide. The five transmembrane helix (a1 - a5) secondary structure model for human ATPase 6 (Cox *et al.*, 1986) predicts that Ser^{99} lies in the NH_3-terminal end of helix a3. Whether this part of helix a3 is in or outside the membrane is still

ambiguous (L. Hatch, pers. comm.). Changes in the loop region between helix a2 and helix a3 in *E. coli* affect the positioning of helix a4 (which is the functionality). Additionally, substitution of endogenous prolines within *E. coli* protein *a* affects the orientation of residues on the helical faces (Howitt *et al.*, 1993). If the same is true for human ATPase 6 (where prolines 89 and 94 are probably analogous to 122 and 144 in *E. coli*), an additional proline at residue 99 will interfere with this orientation. Proline residues, when in the core of alpha helices, produce kinks in the helix structure (von Heijne, 1991). Whilst lipid bilayer membranes are uniquely suited to accommodate such kinks, the introduction of proline into a non-kinked helix will disrupt the structure of the protein, giving rise to the notion that prolines are helix-breakers (Chou and Fasman, 1978). Therefore the introduction of a proline at residue 99 is predicted to affect the ATPase 6 secondary and tertiary structure.

For a mutation to be pathogenic the affected amino acid is likely to be in a highly conserved region of the encoded protein (Riordan-Eva and Harding, 1995). Ser^{99} is part of the hexamer Pro^{94}, Thr, Thr, Gln, Leu, Ser^{99} with 100% conservation in vertebrates (Fig. 2.9). Sea urchins have conserved Ser^{99} although other parts of the domain are less conserved. Invertebrates such as *Drosophila* and honey bees have other residues in this position (see Fig. 2.9 for references). Therefore Ser^{99} could be considered to have medium conservation between species. No species analysed thus far has a proline in this position.

Heteroplasmic mutant	PTTQL**P**MNL
Normal human	PTTQL**S**MNL
Bovine	PTTQL**S**MNL
Armadillo	PTTQL**S**MML
Rat	PTTQL**S**MDL
Whale	PTTQL**S**MNV
Chicken	PTTQL**S**LNM
Ostrich	PTTQL**S**MNM
Xenopus laevis	PTTQL**S**MNM
Salmon	PTTQL**S**LNM
Sea urchin	ATSLI**S**LTY
Honey bee	LTSHL**L**NMI
Drosophila melanogaster	STSHL**T**LTL

Figure 2.9. Between species amino acid homology over the nanapeptide Pro^{94} through Leu^{102}. Sequences were obtained from the following references: Bibb *et al.*, 1981; Anderson *et al.*, 1982; Roe *et al.*, 1985; Cantore *et al.*, 1989; Gadeleta *et al.*, 1989; Thomas and Beckenbach, 1989; Desjardins and Morais, 1990; Arnson *et al.*, 1993; Arnson *et al.*, 1997; Harlid *et al.*, 1997. Alignment is based on homology of the ATPase 6 active site and other highly conserved domains. Residue 99 in humans and homologous residues in other animals are bolded.

Human sperm require ATP for motility. Ejaculated sperm are very flexible in their capacity to obtain energy from other substrates and much ATP is obtained from glycolysis rather than OXPHOS (Ford and Harrison, 1981). However, the addition of rotenone, an OXPHOS inhibitor, to samples of ejaculated semen causes ATP depletion and sperm immobilisation (de Lamirande and Gagnon, 1992). Folgerø *et al.* (1993) have demonstrated this requirement for OXPHOS derived ATP clinically in a patient with a $tRNA^{Leu}$ mutation, where the mutation causes a decrease in motility that could be partially increased by substrate therapy. It is not known whether the mutation found in the present study affects sperm motility, since no mature sperm were found with the mutation in homoplasmy, but the mutation may affect the ability of the sperm to mature fully. The stable nature of this mutation suggests that the mutation must be present in the type A spermatogonia, the earliest precursor germ cells. The role of OXPHOS in spermatogenesis is not known. However, the extra prominent cristae present in sperm mitochondria, and congregation of the mitochondria at the caudal end of the nucleus near the developing flagellum during spermatogenesis, may imply a functional role for the mitochondria in spermatogenesis (Glover *et al.*, 1990).

It is not known if the heteroplasmic mutation at nt 8821 is the cause of the subnormal seminal characteristics seen in semen samples QR568B and QD205F. QF205F has 100% abnormal head morphology in mature sperm that do not have detectable levels of the mutation. The proportion of mutant DNA in PCR products from bulk DNA preparations was 16 to 18%. This is probably too low to be seen as the causal genetic lesion. Unfortunately, the only other study suggesting point mutations having an effect on seminal characteristics (Folgerø *et al.*, 1993) does not quantify the level of heteroplasmy in the individual seminal samples although it is assumed that the point mutation is heteroplasmic (Folgerø *et al.*, 1995). The five other pathogenic mutations seen in the ATPase 6 gene (Introduction, this Chapter) all require above 90% mutant DNA for phenotypic effects within somatic tissues.

Chapter 3

COII GENE MUTATIONS

Introduction

COII structure and function

Cytochrome *c* oxidase subunit II (COII) is one of three mitochondrially encoded subunits of the 13 subunit Cytochrome *c* Oxidase (COX), the terminal electron acceptor in the electron transport chain. COX catalyses the reduction of oxygen to water; a reaction that is coupled to the translocation of protons across the inner mitochondrial membrane. In addition to 13 polypeptides, COX also contains two haem centres and three copper atoms at the active sites (for a review see Capaldi, 1990). The three mitochondrially encoded subunits (COI, II, and III) make up the catalytic core of COX. COII has two transmembrane α helicies (as predicted by hydropathy plots) and a large hydrophilic domain which contains the ligands for Cu_A binding in addition to residues that interact with Cytochrome *c* (Tsukihara *et al.*, 1996). COI and COII interact to provide a catalytic core, whilst COIII may have a role in assembly, stability or activity modulation (Capaldi, 1990). The structures and active sites of all three mitochondrially encoded subunits have been ascertained from bovine heart. Many residues assigned important roles in bovine heart are, however, conserved throughout evolution suggesting that the structures and active sites are similarly conserved (Fig. 3.1).

The amino acid residues in COII thought to be important catalytically include His^{161}, Cys^{196}, Cys^{200}, and His^{204} which are ligands for Cu_A binding, Asp^{112}, Glu^{114}, Asp^{158}, and Glu^{198} which interact with Cytochrome *c* and Tyr^{105}, Trp^{106}, Tyr^{108} and Tyr^{110} which may provide a conduit for electron transfer to haem 1 in COI. Glu^{198} may have an additional role in forming the Cu_A binding site and be a ligand for a regulatory magnesium atom (Capaldi, 1990). All of the above residues fall in the hydrophilic C-terminal arm of COII (Fig. 3.2). In addition to the ligands in the catalytic sites there are a number of amino acid residues that may be important structurally. For example, Asp^{57} and His^{52} are bonded by a hydrogen atom to a phosphotidyl ethanolamine via the imide groups of the main chain on the matrix side of the inner mitochondrial membrane

```
            1111111111222222222233333333334444444444555555
    12345678901234567890123456789012345678901234567890123 45
COW MAYPMQLGFQDATSPIMEELLHFHDHTLMIVFLISSLVLYIISLMLTTKLTHTST
    ::   : : ::::::::::::  :::: ::: :::  ::::   : :::::: :
MAN MAHAAQVGLQDATSPIMEELITFHDHALMIIFLICFLVLYALFLTLTTKLTNTNI
             A                    V                      *

                                                    11111111111
    55556666666666777777777788888888899999999990000000001
    67890123456789012345678901234567890123456789012345678 90
COW MDAQEVETIWTILPAIILILIALPSLRILYMNDEINNPSLTVKTMGHQWYWSYEY
     :::: :: ::::::::: :::::::::::: :: : :::: :  :::::: :::
MAN SDAQEMETVWTILPAIILVLIALPSLRILYMTDEVNDPSLTIKSIGHQWYWTYEY
                                        A          ?

    11111111111111111111111111111111111111111111111111111111
    11111111222222222233333333334444444444555555555566666 6
    12345678901234567890123456789012345678901234567890123 45
COW TDYEDLSFDSYMIPTSELKPGELRLLEVDNRVVLPMEMTIRMLVSSEDVLHSWAV
    :::  : : ::: :   : :: :::: :::::::: :  :::   : :::::::
MAN TDYGGLIFNSYMLPPLFLEPGDLRLLDVDNRVVLPIEAPIRMMITSQDVLHSWAV
     *       P                          T

    11111111111111111111111111111111112222222222222222222 2
    66667777777777888888888899999999990000000000111111111 2
    67890123456789012345678901234567890123456789012345678 90
COW PSLGLKTDAIPGRLNQTTLMSSRPGLYYGQCSEICGSNHSFMPIVLELVPLKYFE
    : ::::::::::::::::    ::: :::::::::: :::::::::: ::: ::
MAN PTLGLKTDAIPGRLNQTTFTATRPGVYYGQCSEICGANHSFMPIVLELIPLKIFE
                         M

     2222222
     2222222
     1234567
COW  KWSASML
          :
MAN  MGPVFTL
      R
```

Figure 3.1. COII polypeptide amino acid alignment between bovines and humans. Numbering denotes amino acid residue. (:) denotes synonymous residues between bovines and humans. * Indicates important functional sites that are not conserved. Underlining indicates important functional sites (Capaldi, 1990) that are conserved. Letters under sequence are human polymorphic amino acids as per (Wallace *et al.*, 1995). ? Indicates site that Capaldi (1990) assigned the residue Trp^{101} and hence is obviously ambiguous.

between helices a1 and a2 (Fig. 3.2). Additionally, Gly^{8}, on the cytosolic side, is bonded by a hydrogen atom via a main chain imide, to a phosphotidyl glycerol which also interacts with residues in COI and a nuclear encoded subunit COVIc (Tsukihara *et al.*, 1995). Lys^{49} does essentially the same thing but on the matrix side. This leads to a complex picture that has the membrane spanning α helicies of COII in a structural role aligning the functional hydrophilic tail for its catalytic role. These functional sites have been ascertained both by high resolution X-ray spectroscopy and mutational studies in yeast and bacteria. Few studies have been performed in humans showing changes that reduce COX activity. No primary pathological mutations have been definitively assigned to the COII gene. A number of population variants have been characterised that change amino acid residues within the COII protein. None of these variants change important ligands, and, in fact, many occur at sites that vary between species (Fig. 3.1).

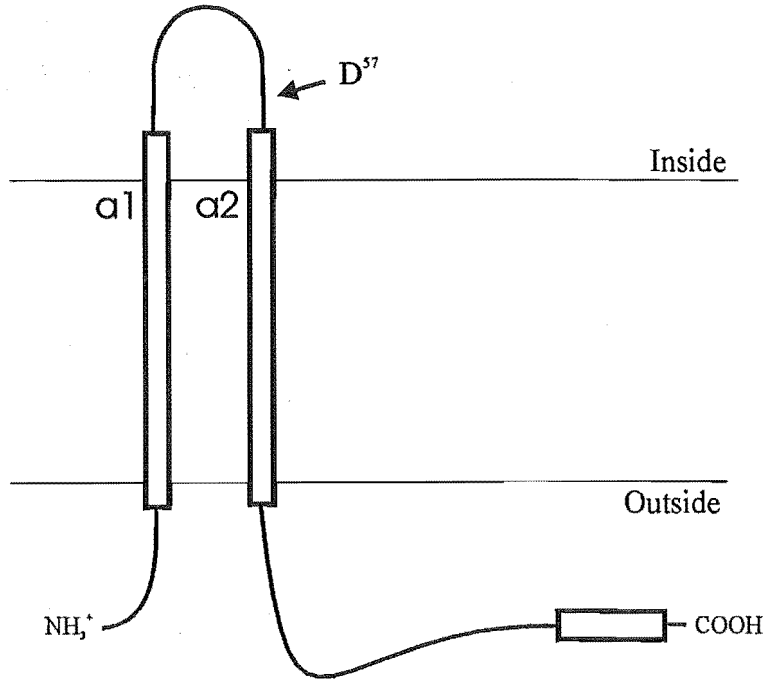


Figure 3.2. Proposed structure of the human COII polypeptide (derived from Tsukihara *et al.*, 1996). Membrane spanning helicies are labelled a1 and a2. Inside and outside, refer to the orientation within the inner mitochondrial membrane. For reference to D^{57} see results.

Population studies based on COII gene variability

Probably 90% of the known COII gene variants are derived from RFLP studies with the highest intensity RFLP studies only screening about 20% of the COII gene.

Additionally, where RFLPs are found, frequently the change causing the loss or gain of the restriction site is not characterised. This allows many different mutations to account for the same RFLP pattern. The mtDNA variation database (MITOMAP) lists 26 out of 676 (3.8%) COII nucleotide positions that are polymorphic (Wallace *et al.*, 1995). Given that most of these changes have been found by RFLP analysis, this is probably a very conservative estimate. Few studies have sequenced the COII genes to the extent that within human variation rates can be assigned. Ruvolo *et al.* (1993) specifically sequenced the COII gene from six individuals (Northern European (Cambridge), Asian and 4 African) and found 7 polymorphic sites within 3 of the individuals when compared to the Cambridge sequence. Horai *et al.* (1992) and Kondo *et al.* (1993) analysed nucleotide and protein sequences (respectively) between primates (with just one human sequence) for 6 mtDNA protein coding genes and found that COII was second only to COI for having the lowest between primate rate of substitution per base. Recently, Davis *et al.* (1997) sequenced the COI, COII and COIII genes from 506 individuals with Alzheimers and 95 controls. In addition to a common 'heteroplasmy', (falsely assigned; see Wallace *et al.* (1997), Hirano *et al.* (1997), and Chapter 4) they presumably also found a large number of polymorphisms. This polymorphic data is yet to be published. This would be the most definitive study yet of COII gene variability between humans.

In the present study the COII amplicon includes three tRNA genes and a 25 bp intergenic region. Both $tRNA^{Lys}$ and $tRNA^{Ser(UCN)}$ can harbour primary pathogenic mutations that cause diseases (e.g. Reid *et al.*, 1994; Nakamura *et al.*, 1995). Pathogenic mutations in these genes can be heteroplasmic or homoplasmic; heteroplasmic being the more common as homoplasmic changes in tRNA genes are often lethal. Where these changes are relevant to the present study they are discussed. The 25 bp intergenic spacer between COII and $tRNA^{Lys}$ has no known function. That it is conserved, at least in part, in all species studied so far suggests that it may have an undescribed role. Analysis of post-transcriptionally modified mRNA suggests that the COII mRNA (transcribed off the mtDNA heavy strand) ends with a polyA tail 3' of the intergenic spacer suggesting that nucleotides within this spacer serve as a polyadenylation signal (Ojala *et al.*, 1981). Within this spacer there is a 9 bp direct repeat, one copy of which is often missing in Asian, Polynesian and African populations. This and other changes, where relevant, are discussed below.

In line with the previous chapter, the aims of the present study were to assess the variability in the COII gene and identify mutations in this gene in a Christchurch male population and to determine if these mutations can be related to the fertility status of the individuals carrying the mutations.

The COII gene was chosen for this analysis because it has one of the lowest levels of intraspecific variation in humans (Marzuki *et al.*, 1991) probably due to its structural and functional constraints. Additionally, the COII polypeptide, like the ATPase 6 polypeptide is well characterised, allowing predictions about the effect gene mutations have on protein structure and function.

Materials and Methods

The rational and method of screening seminal samples for changes in the COII gene are essentially the same as for the ATPase genes (Chapter 2).

Sample statistics

Two hundred and twenty three semen samples were screened for changes in the COII amplicon with an average count of 72 x 10^6 sperm/ml (± 78), a motility of 48% (± 20) with the average age of donor of 36 years (to year of sample donation). Many of these semen samples were used for ATPase gene screening as described in the previous chapter (see Appendix 1).

Polymerase chain reaction

COII primers were chosen using the OSP computer programme in such a way that they did not fall on common polymorphic sites, did not anneal to themselves or each other with high affinity, and had melting points above 50°C for high specificity binding. The primers chosen were HMTL712 (5' CCTACGCCAAAATCCATTTC) and HMTH844 (5' agttgggtgatgaggaatag). The COII amplicon, consisting of 1316 bp, encompasses 319 bp of COI, $tRNA^{Ser(UCN)}$, $tRNA^{Asp}$, COII, a 25 bp intergenic spacer, $tRNA^{Lys}$ and 76 bp of ATPase 8. PCR conditions for the COII amplicon were

optimised for $MgCl_2$ concentration and annealing temperature. Optimal cycling conditions were an initial five cycles of 94°C/30 sec, 60°C/30 sec, and 72°C/1 min 20 sec followed by a further five cycles of 94°C/30 sec, 56°C/30 sec, and 72°C/1 min 20 sec. This was followed by 25 cycles of 94°C/30 sec, 54°C/30 sec, and 72°C/1 min 20 sec. The 25 µl PCR reaction mixtures contained 5 µl total cell DNA prepared as in the previous chapter, 20 pmol of each primer, 5 nmol of each dNTP, 2.5 µl of 10 x reaction buffer (500 mM KCl, 100 mM Tris-HCl (pH 9.0)), 1.4 mM $MgCl_2$ and 1 unit of *Taq* polymerase (Promega or Boehringer Mannheim).

Mutation detection

A 10 µl aliquot of the PCR product was digested to completion with two units of *Bst*N1 (New England Biolabs) in the appropriate buffer in a final volume of 20 µl at 60°C. *Bst*N1 cleaves the COII amplicon twice to give fragments of 597 bp, 467 bp and 252 bp. SSCP screening was carried out on the *Bst*N1 cleaved COII amplicon as described in the preceding chapter. *Bst*N1 occasionally gives additional bands as a result of nonspecific activity; however, these are larger than 597 bp (the largest *Bst*N1 fragment), and so did not hinder mutation detection. Undenatured DNA when electrophoresed on SSCP gels, gave single bands that did not interfere with scoring of denatured bands. When changes in the SSCP pattern were observed, the screening was repeated for confirmation. Once different patterns were confirmed, the COII PCR product was cleaved with both *Hae*III and *Acc*1 restriction enzymes and then re-analysed by SSCP analysis. The DNA fragments carrying the nucleotide changes were identified by comparing the mobilities of the fragment bands with those of the controls. The digested fragments containing the mutations were cloned and sequenced.

Cloning and colony selection for mutation characterisation

Competent *E. coli* DH5α were prepared from an existing laboratory *E. coli* DH5α glycerol preparation. A single colony was selected from a streaked *E. coli* DH5α LB plate and grown overnight at 37°C, shaking at 250 rpm in 10 ml of LB broth. Two and a half millilitres of this culture was added to 100 ml LB broth and grown, shaking as

before, at 37°C for 2 hours and 40 min (A_{550} between 0.6 and 0.7). Cells were collected by centrifugation at 3,000 x g for 5 min at 4°C in a Sorvall centrifuge. The cell pellet was resuspended in 15 ml of ice cold 50 mM $CaCl_2$, 5 mM Tris (pH 8.0) using a sterile siliconized pasteur pipette before being left to sit on ice for 20 min. Cells were again pelleted as before, and then re-suspended in 2 ml of ice-cold 50 mM $CaCl_2$, 5 mM Tris (pH 8.0), 15% (v/v) glycerol. Two-hundred microlitre aliquots were transferred to prechilled tubes and kept at 4°C for 19 hours. Cells were then snap frozen in liquid N_2 before being stored long term at -80°C.

The cloning vector used in this study is the phagmid pBluescribe M13+ (pBSM13+) (Stratagene). pBSM13+ is a multicopy phagmid (500 - 700 copies) derived from pUC with an M13 origin of replication and binding sites for both T7 and T3 promoter primers (T7 p/p and T3 p/p). Additionally, pBSM13+ contains the selectible Ampicilin (Amp) resistance gene conferring Amp resistance to host bacteria and the LacZ gene, in which the multiple cloning site (MCS; pUC 19 polylinker) is found, allowing colour selection of recombinant colonies. Native pBSM13+ is 3204bp long. pBSM13+ was prepared from pBSM13+ transformed *E. coli* DH5α. Single colonies were selected from an LB + Amp (50 μg/ml) plate streaked with *E. coli* DH5α :: pBSM13+ and grown overnight in 10 ml of LB + Amp (50 μg/ml) broth at 37°C shaking at 250 rpm. pBSM13+ was purified from these overnight cultures using a modified alkaline lysis and PEG precipitation protocol (Applied Biosystems Inc.). DNA concentrations and purity were assessed by spectrophotometric readings at A_{260} and A_{280} in an LKB ultraspec II spectrophotometer. Purified phagmid was cleaved with *Sma*1 (Boehringer Mannheim) at 25°C overnight. A small sample of the digest was checked for complete digestion by agarose gel electrophoresis. Linearized phagmid DNA was precipitated with 0.1 volume of 3 M sodium acetate (pH 7.0) and 2 volumes of 100% ethanol at -20°C for 15 min before being pelleted by centrifuging at 12,000 x g at 4°C for 10 min. The pellet was washed with 70% ethanol, spun as before and dried for 3 min under vacuum to remove traces of ethanol. Purified DNA was re-suspended in 90 μl of 10 mM Tris-HCl (pH 8.3) and 10 μl 10 x calf intestinal alkaline phosphatase (CIAP) buffer (Sambrook *et al.*, 1989). Three units of CIAP were added and the mix was incubated at 37°C for 20 min before the addition of another 3 units of CIAP with

subsequent incubation at 55°C for a further 40 min. This was then heat inactivated at 75°C for 10 min. The de-phosphorylated phagmid DNA was then cleaned using standard phenol/chloroform extraction procedures. The purified DNA was dissolved in TE8 prior to ligation.

Fragments to be cloned were amplified in a 100 µl PCR mix (the same as a standard 25µl mix but scaled up 4 times). Fifty microlitres of the PCR products was digested with the appropriate restriction enzyme(s) overnight. As none of the restriction sites within the COII amplicon were compatible with those of the multiple cloning site in pBSM13+, all the fragments had to be blunt end ligated. Between one and two units of Klenow polymerase (Boehringer Mannheim) was used to blunt end the PCR product digests at 37°C for 30 min. The reaction was inactivated at 75°C for 10 min. Digests were electrophoresed in a 3% agarose gel and the bands of interest were extracted by either a 'freeze and squeeze' protocol (Towner, 1993) or using a commercial gel extraction kit (BioRad or Boehringer Mannheim). The DNA obtained by the extraction kits was sufficiently clean for cloning, whereas the DNA obtained by the 'freeze and squeeze' extraction method required purification. One volume of 4 M ammonium acetate and two volumes of isopropanol were added to the DNA fragments recovered from agarose gels and kept at room temperature for 15 min and then the DNA was pelleted by centrifugation at 12,000 x g for 10 min at room temperature. The DNA was then washed in 70% ethanol, vacuum dried and re-suspended in TE8.

Ligations were carried out essentially as per Sambrook *et al.* (1989) with a few modifications. The highest efficiency of blunt-end ligation was found to be with an insert to vector molar ratio of between 6:1 and 12:1. In most instances between 50 and 75 ng of linear and de-phosphorylated vector DNA was ligated. Seven microlitres of vector and insert mix (in a 1:6 molar ratio) were heated to 65°C for 2 min then cooled on ice. One microlitre of 10 x DNA ligation buffer (200 mM Tris-HCl (pH 7.6), 50 mM $MgCl_2$, 50 mM DTT, 500 µg/ml BSA Fraction V), one microlitre 10 mM ATP and one unit (1 µl) T4 DNA ligase (Gibco BRL) was then added to the DNA mix. Ligations were carried out at 16°C for between 13 and 16 hours.

Either 5 or 10 µl of the ligation mix was used to transform 200 µl of competent cells. The DNA was added, briefly mixed and then left on ice for a further 40 min. Cells were heat shocked for 2 min at 42°C, then place on ice for 2 min before the addition of 400 µl SOC media (Appendix 2). Tubes were gently inverted and incubated at 37°C for 1 hour. One hundred, 200 and 300 µl aliquots of transformed cells were plated on prewarmed SOC + Amp + Xgal + IPTG plates, dried and allowed to grow discrete colonies by incubating for 24 hours inverted at 37°C. For transformant colour selection, plates were placed at 4°C for at least 48 hours before being scored against a white background. Appropriate control transformations were performed to check for spontaneous Amp reversion. If subsequent colonies were still to small for accurate colour assessment, but plates were becoming over grown, 100 colonies were randomly selected and spotted in a regular array on SOC + Amp + Xgal + IPTG plates, and then grown as above.

Separation of true and false positive recombinants was done using an alkaline lysis preparation. Colonies were transferred by toothpick into 3 ml LB + Amp (50 µg/ml) in sterile McCartney bottles. Cultures were grown, shaking at 250 rpm, with lids loosely capped, for 12 hours at 37°C. In addition to inoculating a culture, the colony was also streaked on an LB plate with the same toothpick and grown at 37°C overnight to provide a culture backup. One and a half millilitres of culture was centrifuged at 12,000 x g at 4°C for 30 sec to obtain a cell pellet and the supernatant was aspirated off. The cell pellet was re-suspended in 350 µl of 0.1 M NaCl, 10 mM Tris-HCl (pH8.0), 1 mM EDTA, 5% (v/v) Triton X-100 (STET), and 25 µl of 10 mg/ml lysozyme (Boehringer Mannheim) in 10 mM Tris (pH 8.0) by vortexing for 3 sec. Suspensions were boiled for 40 sec before pelleting the bacterial debris by centrifugation at 12,000 x g for between 5 and 10 min. Debris were removed with a sterile toothpick before adding 40 µl of 3 M sodium acetate (pH 5.2) and 420 µl of isopropanol. To aid precipitation this was vortexed and left to stand at room temperature for 5 min. Pelleting, washing and drying were performed as above with the recombinant phagmid pellet being re-suspended in TE8 + RNase A (20 µg/ml) (Boehringer Mannheim). The small insert size meant that resolving recombinant and non-recombinant uncut phagmids on an agarose gel was difficult. Therefore phagmids were cleaved with *Pvu*II (Boehringer Mannheim) which cuts on either side of

the MCS at nt 766 and nt 1148 (and not in the COII amplicon) and then these cleavage products were resolved on 2% agarose gels. *Pvu*II digestion of non-recombinant phagmids yields a 382 bp fragment with recombinants having larger fragments.

When large numbers of clones were being analysed from many concurrent cloning reactions alkaline lysis preparations as a means of differentiating recombinants from non-recombinants was inefficient. In these instances colonies were lysed on the plate, transferred to nylon membranes and probed with an α^{32}P dCTP labelled COII PCR product to select for positive colonies as described below. Replica LB + Amp plates were prepared of white colonies and grown for 24 hours at 37°C. Included in each of these plates were blue colonies as a negative control, and known recombinants as a positive control. Hybond N nylon membrane was placed on colonies and left for 1 min to transfer the colony on to the membrane. Membranes were then placed colony side up on a stack of three filter papers soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH) and left for 7 min. The membranes were transferred to a stack of three filter papers soaked in neutralising solution (1.5 M NaCl, 0.5 M Tris (pH7.2), 1 mM EDTA) for 3 min. After repeating the neutralising step the membranes were washed in 2 x SSC then transferred to filter paper to air dry. Membranes were wrapped in Saran wrap and U.V. fixed for 3 min before prehybridising for 1 hour at 65°C in prehybridisation solution (Appendix 2). The DNA probe was prepared by boiling 100 ng of COII PCR product for 10 min then quenching in an ice slurry (this DNA had been prepared by ammonium acetate / isopropanol precipitation). A random prime reaction containing 2 µl (100 ng) DNA, 3 µl dATP, dGTP, dTTP, 2 µl reaction mix (Boehringer Mannheim), 4 µl 50 µCi α^{32}P dCTP, 8 µl ddH_2O and 1 unit (1 µl) Klenow polymerase was prepared and incubated at 37°C for 30 min. This reaction was stopped by heat inactivating at 65°C for 10 min. The probe was boiled for 8 min, diluted with pre-heated prehybridisation solution and then added to membranes. Hybridisation was carried out overnight at 65°C, in a shaking waterbath. After hybridisation, membranes were washed twice with pre-heated 0.3 M NaCl, 0.3 M sodium citrate (2 x SSC) for 15 min at 65°C with shaking, once with pre-heated 2 x SSC, 0.1% (v/v) sodium dodecyl sulphate (SDS) for 15 min at 65°C with shaking and once with pre-heated 0.1 x SSC, 0.1% (v/v) SDS for 15 min at

65°C with shaking. Membranes were air dried, wrapped in Saran wrap and exposed to X-ray film (Amersham Hyperfilm MP) for between 5 and 24 hours.

COII heteroplasmy analysis

When heteroplasmic SSCP patterns were identified and confirmed, the locations of the mutation(s) were identified to short sequences by digestion with restriction enzymes. COII heteroplasmies were analysed by cloning using pBSM13+. COII digestion fragments with heteroplasmic SSCP patterns were blunt end ligated into blunt ended *Sma*1 cut pBSM13+. pBSM13+ contains a high copy number pMB1 replicon (Sambrook *et al.*, 1989). To obtain only one clone type (normal or mutant) per colony, low efficiency transformation was performed by using small amounts of ligated DNA and large numbers of competent cells. Because inserts were blunt end ligated, each of the mutant or normal clones contained inserts in one of two orientations. SSCP was used in clone analysis. *Pvu*II digests of recombinant colony boiling preparations were electrophoresed on 5% SSCP gels using the same conditions as outlined. A minimum of four patterns were seen on these SSCP gels (normal and reverse orientations had different patterns because the *Pvu*II digests contained invariant phagmid DNA in addition to the insert). The orientation of each insert was ascertained by cutting the positive phagmids with enzymes that cleaved both in the insert and vector (different orientations giving different sized cleavage fragments). In this way, the SSCP patterns were explained by only the presence of normal and mutant DNA. Assuming that both normal and mutant fragments were cloned with equal efficiency, the proportion of clones with a mutant pattern was used to ascribe the proportion of mutant fragments in the PCR mixture. Representatives of both normal and mutant SSCP patterns were sequenced from duplicate cloning reactions using the T7 sequencing reaction.

Sequencing for mutation characterisation

DNA obtained by alkaline lysis preparations of recombinant phagmids was sequenced using a T7 sequencing kit (Pharmacia Biotech). Sequencing gels were prepared and run as per Chapter 2.

Results

Eleven different SSCP patterns were seen from the screening of the *Bst*N1 cleaved COII amplicon from 223 semen samples (Table 3.1.). The most common pattern (not shown in Table 3.1) was assumed to represent the Cambridge sequence as this has a European origin. Of the 223 samples screened, 28% are oligozoospermic 62% of these being severely oligozoospermic (excluding missing data). 46% of samples are asthenozoospermic.

Table 3.1 Confirmed different SSCP patterns after *Bst*N1 screening of COII amplicon from 223 semen samples. As in Table 2.1, the most common pattern, assumed to represent the Cambridge sequence, has been omitted. Patterns have been ascribed retrospectively to aid reader understanding. * = nat and denat refer to native and denatured DNA respectively.

Pattern	Sample	Fragment	Change*
1	JN	467	bands converge
2	95-37	467	nat and denat dropped
2	95-79	467	nat and denat dropped
2	AM436E	467	nat and denat dropped
2	AT665C	467	nat and denat dropped
2	BT	467	nat and denat dropped
2	QN182W	467	nat and denat dropped
2	SF95-160	467	nat and denat dropped
3	95-13	467	top band raised
3	95-24	467	top band raised
3	95-27	467	top band raised
3	AM470K	467	top band raised
3	QD747V	467	top band raised
3	QR509H	467	top band raised
3	QU380K	467	top band raised
3	SF94-299	467	top band raised
3	SF94-409	467	top band raised
3	SF95-122	467	top band raised
3	SF95/612	467	top band raised
4	QD788L	597	top band lowered
5	SF96-279	597	top band lowered
6	CCR	597	top band lowered
7	AM470K	597	top band lowered
8	AM433B	597	heteroplasmy
9	94-107	597	heteroplasmy
10	TM	467	467 top band raised

Seven of the COII SSCP patterns were due to homoplasmic point mutations (Table 3.2). Two of these patterns were found in 7 and 11 out of 223 semen samples. Twenty eight percent (62/223) of the sampled population were oligozoospermic, while two of the six (33%) semen samples with pattern 2 were oligozoospermic (excluding missing data). These proportions are not significantly different (exact binomial probabilities, p = 0.3172, two tailed test with confidence limits of 95%). However, all of the five samples with pattern 2, for which motility was available, were asthenozoospermic. This is significantly different to the 46% of asthenozoospermic samples in the sampled population (exact binomial probabilities, p = 0.02119 *, two tailed test with confidence limits of 95%). Out of 10 samples showing pattern 3 for which data was available, two were oligozoospermic and two were asthenozoospermic. These proportions were not significantly different to the proportions of oligozoospermic and asthenozoospermic semen samples in the sampled population (oligozoospermic; exact binomial probabilities, p = 0.2523, two tailed test with confidence limits of 95%, asthenozoospermic; exact binomial probabilities, p = 0.0669, two tailed test with confidence limits of 95%).

The seven homoplasmic COII SSCP patterns are shown in Figures 3.3 to 3.9. These patterns were analysed as described in their respective figure legends.

Table 3.2. Summary of the homoplasmic changes sequenced within the COII amplicon. (nd) refers to digests not performed. For further information about how these patterns were analysed, refer to the respective pattern figure legends (Fig.).

Pattern	Sample	*Bst*N1 fragment	*Hae*III fragment	*Acc*1 fragment	Change	Position	AA change	Fig.
1	JN	467	364	548	T→C	8077	Ala→Ala	3.3
2	7 samples	467	141	nd	9bp del	8272-8280?	-	3.4
3	11 samples	467	364/141	nd	G→A	8251	Gly→Gly	3.5
4	QD788L	597	300	180	C→T	7476	$tRNA^{Ser(UCN)}$	3.6
			389	548	G→A	7789	Leu→Leu	
5	SF96-279	597	389	548	C→A	7819	Leu→Leu	3.7
6	CCR	597	389	548	G→A	7853	Val→Ile	3.8
7	AM470K	597	389	548	G→A	7754	Asp→Asn	3.9
		467	364/141	nd	G→A	8251	Gly→Gly	

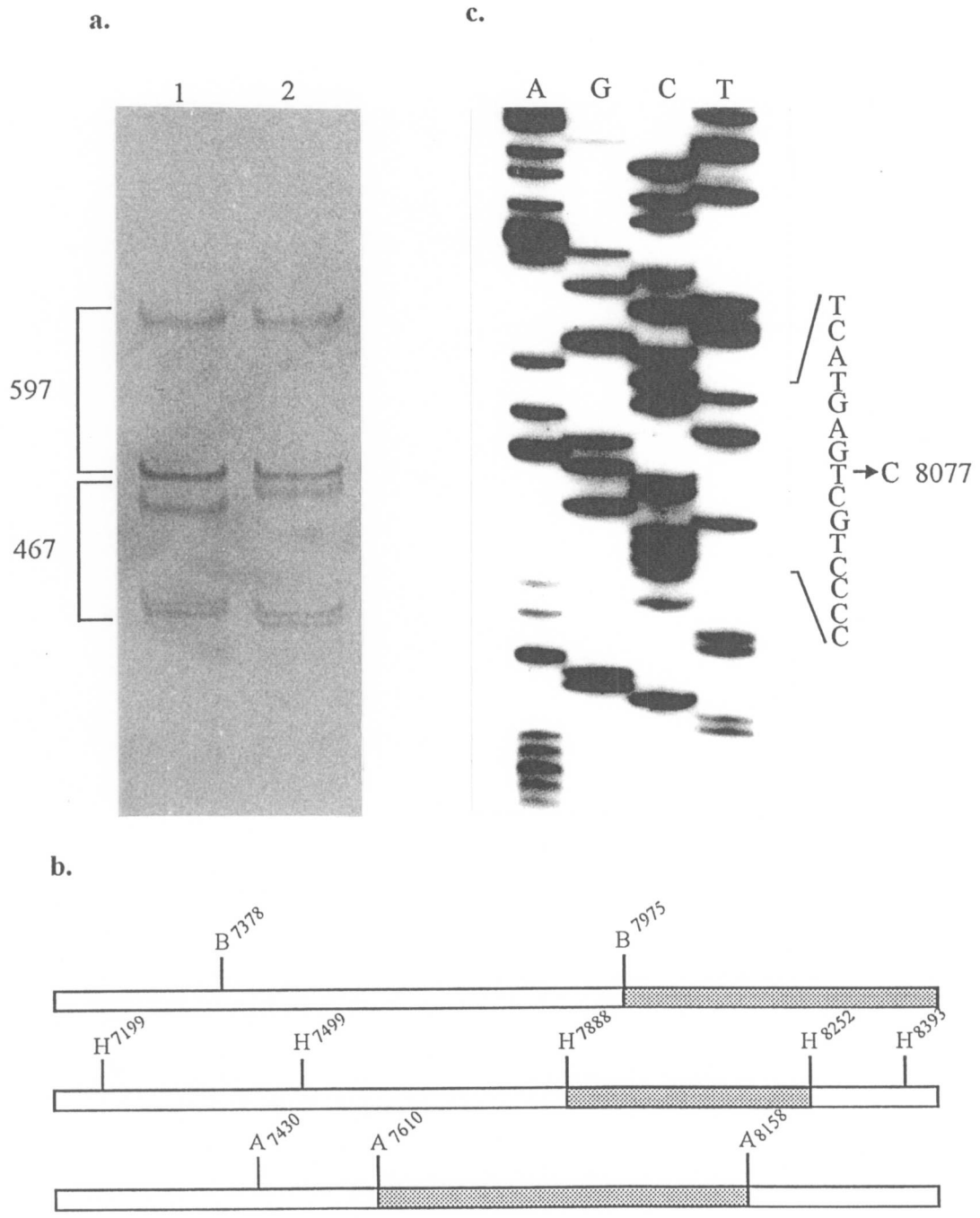


Figure 3.3 **a**. Pattern 1 from COII amplicon screening of semen sample JN. Lane 1 shows the two bands representing the 467 bp fragment from *Bst*N1 cleaved JN COII amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). **b**. Schematic representation of restriction endonuclease cleavage sites within the COII amplicon for enzymes used in the analysis of pattern 1. Restriction sites: *Bst*N1 (B), *Hae*III (H), *Acc*1 (A). Fragments showing mobility difference with respect to control samples are shaded. A 183 bp *Bst*N1/*Acc*1 fragment between nt 7975 and nt 8158 was cloned into *Sma*1 cut pBSM13+. Recombinant clones were selected after two rounds of colony hybridisation with whole COII amplicon. While cloning the 183 bp *Bst*N1/*Acc*1 fragment, a 180 bp fragment was also cloned (due to the proximity to the 183 bp band in the agarose gel). To select just the 183 bp band thought to contain the homoplasmic change for sequencing, all 13 positive clones were cleaved with *Csp*61. Clones containing the 183 bp insert yielded 4 fragments, whilst those with the incorrect 180 bp insert yielded only 3 (*Csp*61 cuts within the 183 bp but not the 180 bp fragment). Ten clones contained the correct insert. Two of the clones (8-1-4 and 7-1-6) were sequenced. **c**. Portion of sequencing gel of clone 8-1-4 showing the T → C transition at nt 8077. This was the only change found and was found in both of the sequenced clones.

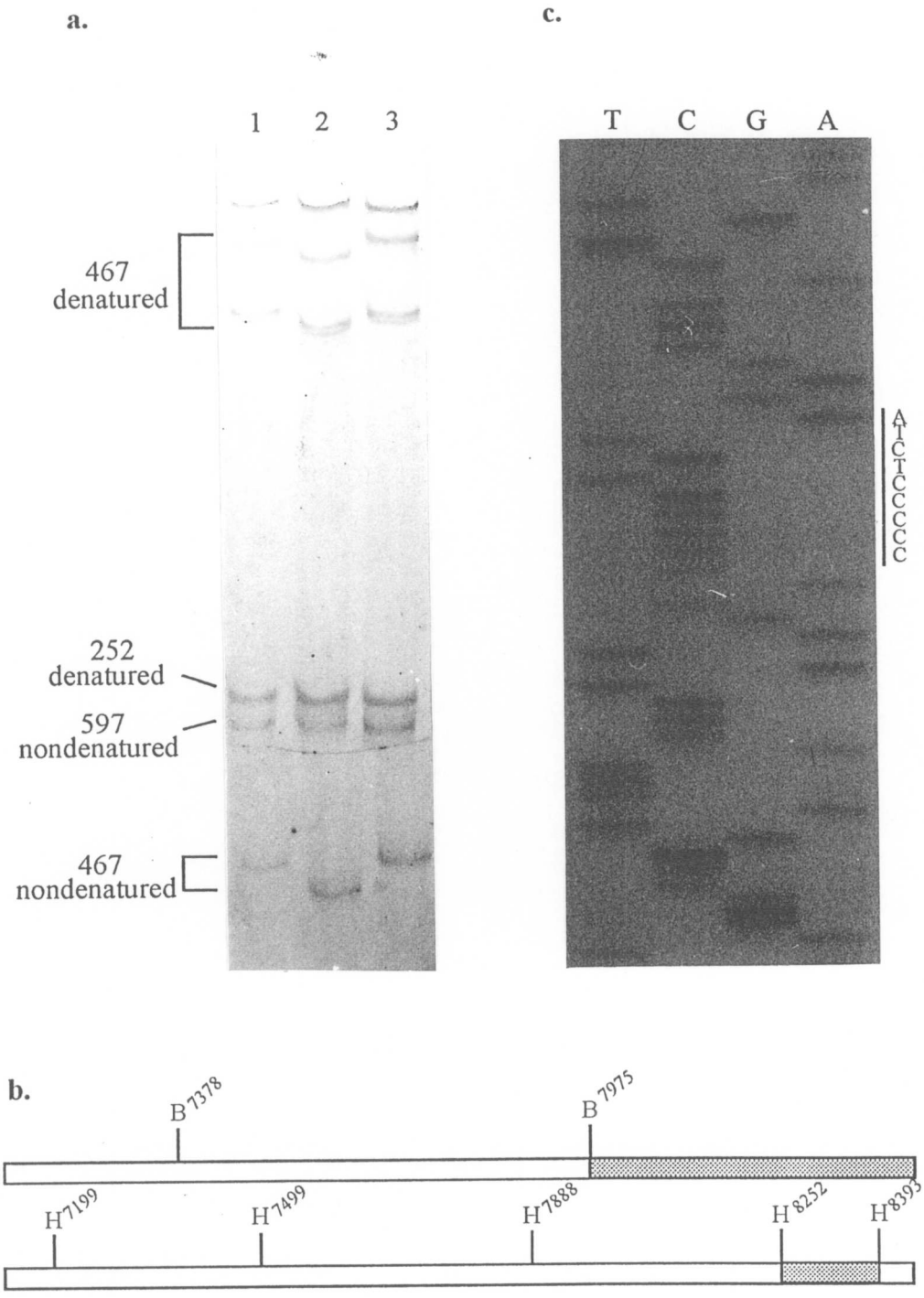


Figure 3.4 **a**. Pattern 2 from COII amplicon screening of semen sample AT665C. Lane 2 shows the two bands representing the 467 bp fragment from *Bst*N1 cleaved AT665C COII amplicon as visualised on a 5% SSCP gel with an increased mobility when, compared with lanes 1 and 3, from control samples with the most common pattern (assumed to have the Cambridge sequence). Note that the 467 bp undenatured band also has a faster mobility in lane 2 when compared with lanes 1 and 3. **b**. Schematic representation of restriction endonuclease cleavage sites within the COII amplicon for enzymes used in the analysis of pattern 2. Restriction sites: *Bst*N1 (B), *Hae*III (H). Fragments showing mobility difference with respect to control samples are shaded. When affected COII amplicons are cleaved with *Hae*III, the deletion was localised to the 141 bp *Hae*III fragment between nt 8252 and nt 8394. Due to the proximity of primers HMTL817 and HMTH844, the COII amplicons of semen samples 95-37, 95-79, QN182W and BT were sequenced directly using cycle sequencing. All sequences showed a 9 bp deletion in the intergenic region between COII and $tRNA^{Lys}$. **c**. Portion of cycle sequencing gel of sample BT primed from HMTH844 showing the lack of one of the 9 bp repeats characteristic of pattern 2 (compare with Figure 3.5**c**)

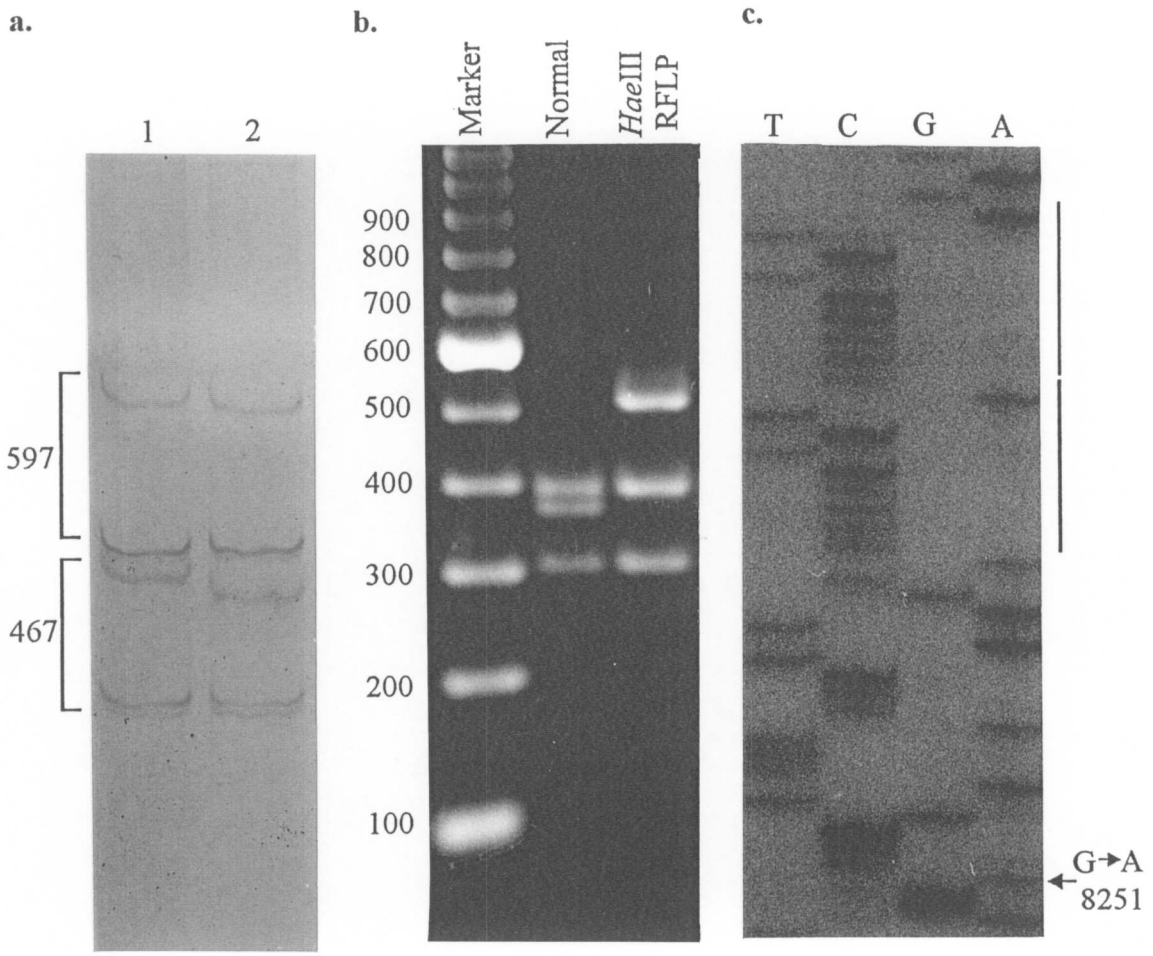


Figure 3.5 **a**. Pattern 3 from COII amplicon screening of semen sample QU380K and 10 other semen samples. Lane 1 shows the two bands representing the 467 bp fragment from *Bst*N1 cleaved QU380K COII amplicon as visualised on a 5% SSCP gel with an increased mobility when compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). When the COII amplicon from these samples was cleaved with *Hae*III it was observed that the 141 bp and 364 bp *Hae*III fragments have been replaced with a fragment of 505 bp. **b**. 3% agarose gel showing pattern 3 *Hae*III RFLP compared with the normal *Hae*III cleavage pattern for the COII amplicon. Marker is a 100 bp ladder (Gibco BRL). Due to the proximity of primers HMTL817 and HMTH844, the COII amplicon of sample QU380K was sequenced directly using cycle sequencing. **c**. Portion of cycle sequencing gel of sample QU380K primed from HMTH844 showing the G → A transition at nt 8251. (compare with Figure 3.4**c**, bars on Figure 3.5**c** represent two copies of 9 bp repeat). All of the samples with pattern 3 were cleaved with *Hae*III and showed the presence of this RFLP. It was assumed that the mutation was the same in all.

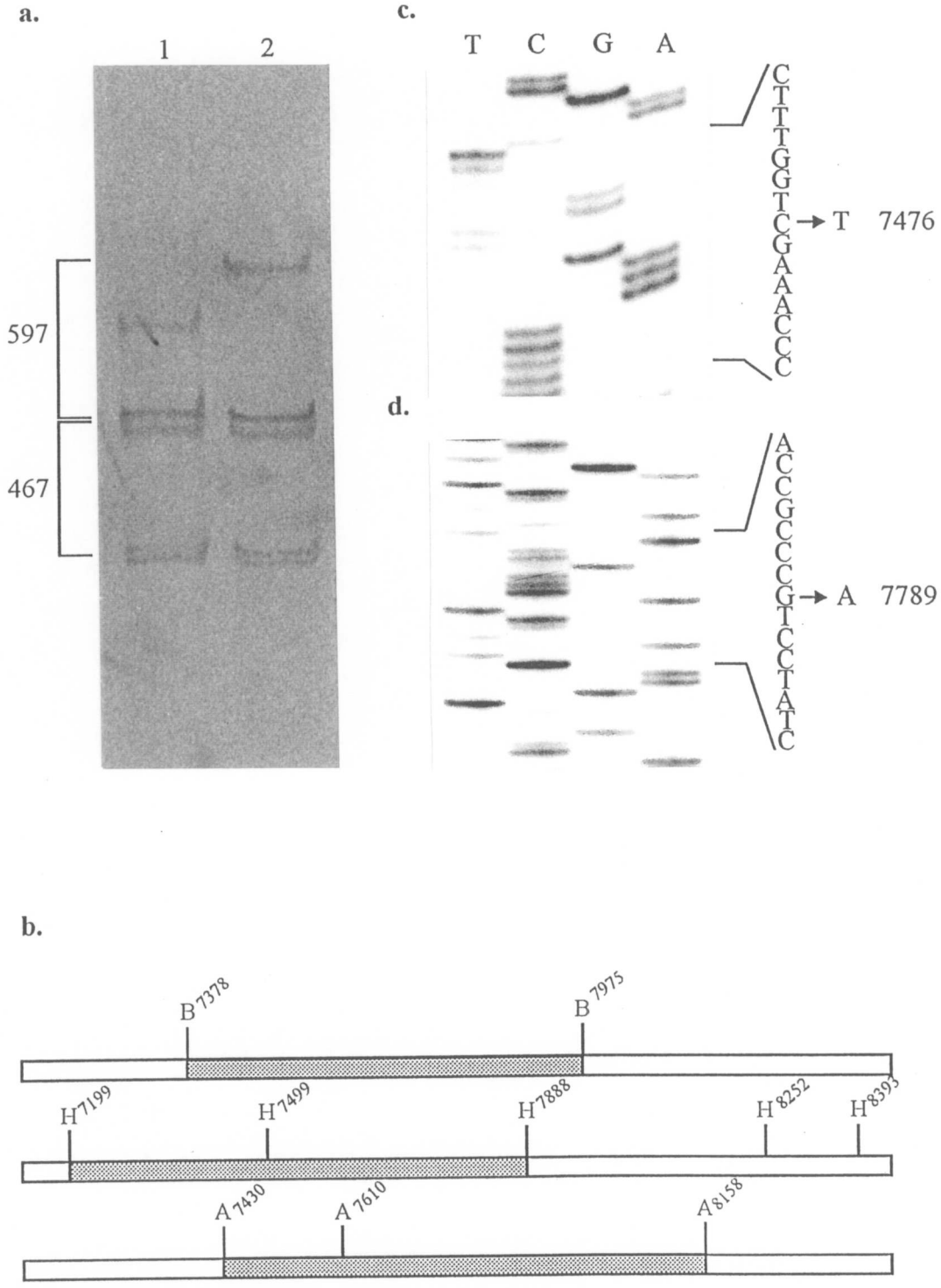


Figure 3.6 **a**. Pattern 4 from COII amplicon screening of semen sample QD788L. Lane 1 shows the two bands representing the 597 bp fragment from *Bst*N1 cleaved QD788L COII amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). **b**. Schematic representation of restriction endonuclease cleavage sites within the COII amplicon for enzymes used in the analysis of patterns 4 and 8. Restriction sites: *Bst*N1 (B), *Hae*III (H), *Acc*1 (A). Fragments showing mobility difference with respect to control samples are shaded. When the COII amplicon of QD788L was cleaved with *Hae*III, bands from the 300 bp and 389 bp fragments had mobility shifts, and when cleaved with *Acc*1, bands from the 180 bp and 548 bp fragments had mobility shifts indicating that there were at least two changes within the QD788L COII amplicon. Clones 5-1-6 and 147-1-3 containing a 365 bp *Acc*1/*Bst*N1 fragment and clones 147-1-2 and 147-2-7 containing a 180 bp *Acc*1 fragment were sequenced. **c**. Portion of sequencing gel of clone 147-1-3 showing the G → A transition at nt 7789. **d**. Portion of sequencing gel of clone 147-1-2 showing the C → T transition at nt 7476.

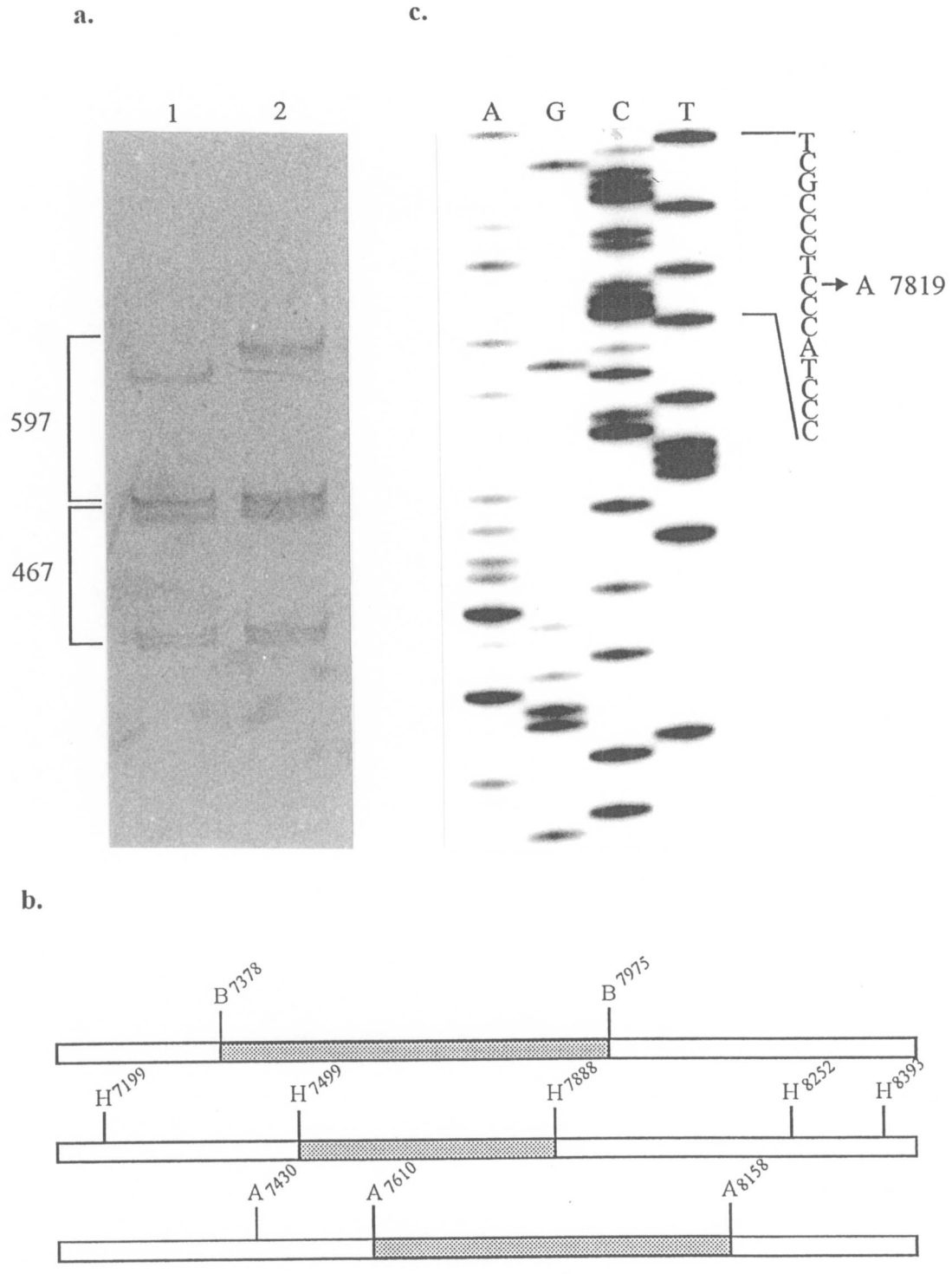


Figure 3.7 **a**. Pattern 5 from COII amplicon screening of semen sample SF96-279. Lane 1 shows the two bands representing the 597 bp fragment from *Bst*N1 cleaved SF96-279 COII amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). **b**. Schematic representation of restriction endonuclease cleavage sites within the COII amplicon for enzymes used in the analysis of patterns 5, 6 and 7. Restriction sites: *Bst*N1 (B), *Hae*III (H), *Acc*1 (A). Fragments showing mobility difference with respect to control samples are shaded. The changes causing pattern 5 were in a 278 bp *Acc*1/*Hae*III fragment between nt 7610 and nt 7888. A 389 bp *Hae*III fragment (nt 7499 to nt 7888) was cloned containing the 278 bp *Acc*1/*Hae*III fragment. Positive clones 165-1-10 and 165-2-5 were sequenced. **c**. Portion of sequencing gel of clone 165-2-5 showing the C → A transversion at nt 7819. (Compare with Figure 3.8**b**.)

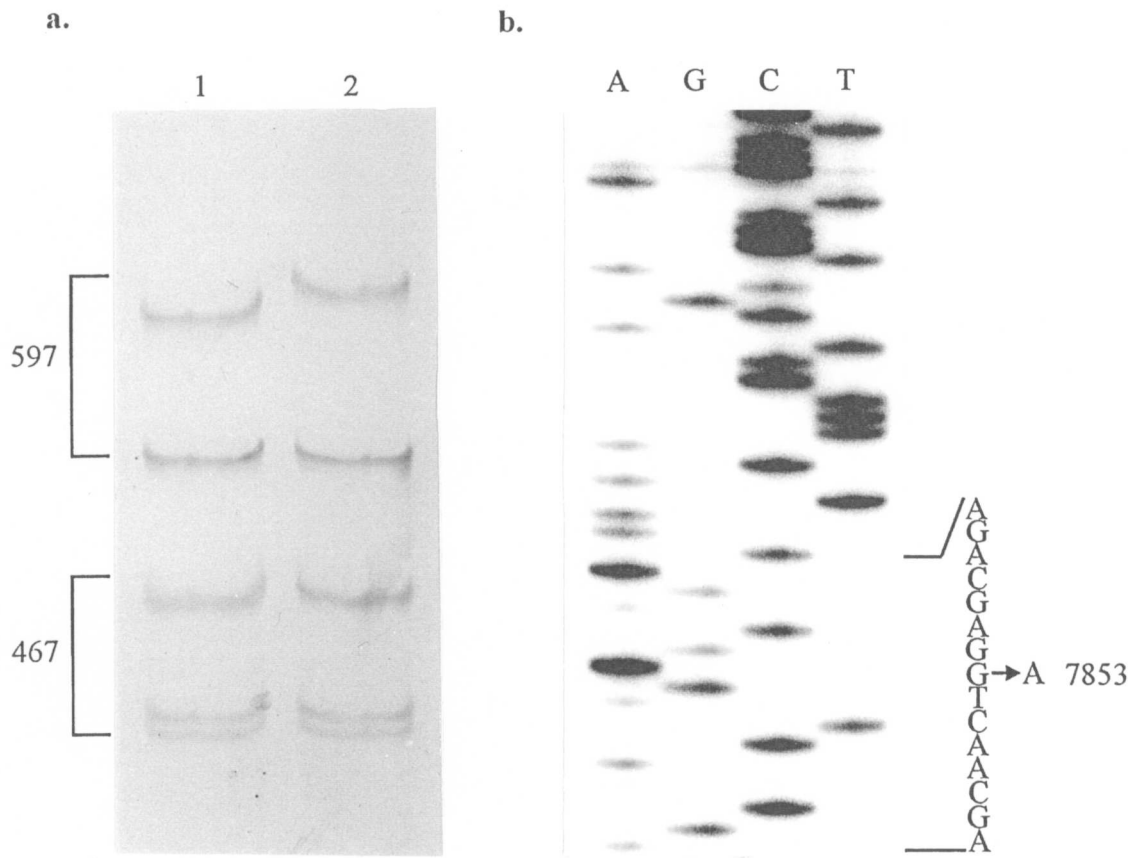


Figure 3.8 **a**. Pattern 6 from COII amplicon screening of semen sample CCR. Lane 1 shows the two bands representing the 597 bp fragment from *Bst*N1 cleaved CCR COII amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). Restriction fragments containing mobility shifts are the same as pattern 5 (Fig. 3.7**b**). The changes causing patterns 6 were therefore in a 278 bp *Acc*1/*Hae*III fragment between nt 7610 and nt 7888. A 389 bp *Hae*III fragment (nt 7499 to nt 7888) was cloned containing the 278 bp *Acc*1/*Hae*III fragment. Positive clones 165-3-8 and 165-4-6 were sequenced. **b**. Portion of sequencing gel of clone 165-4-6 showing the G → A transition at nt 7853. (Compare with Figure 3.7**c**.)

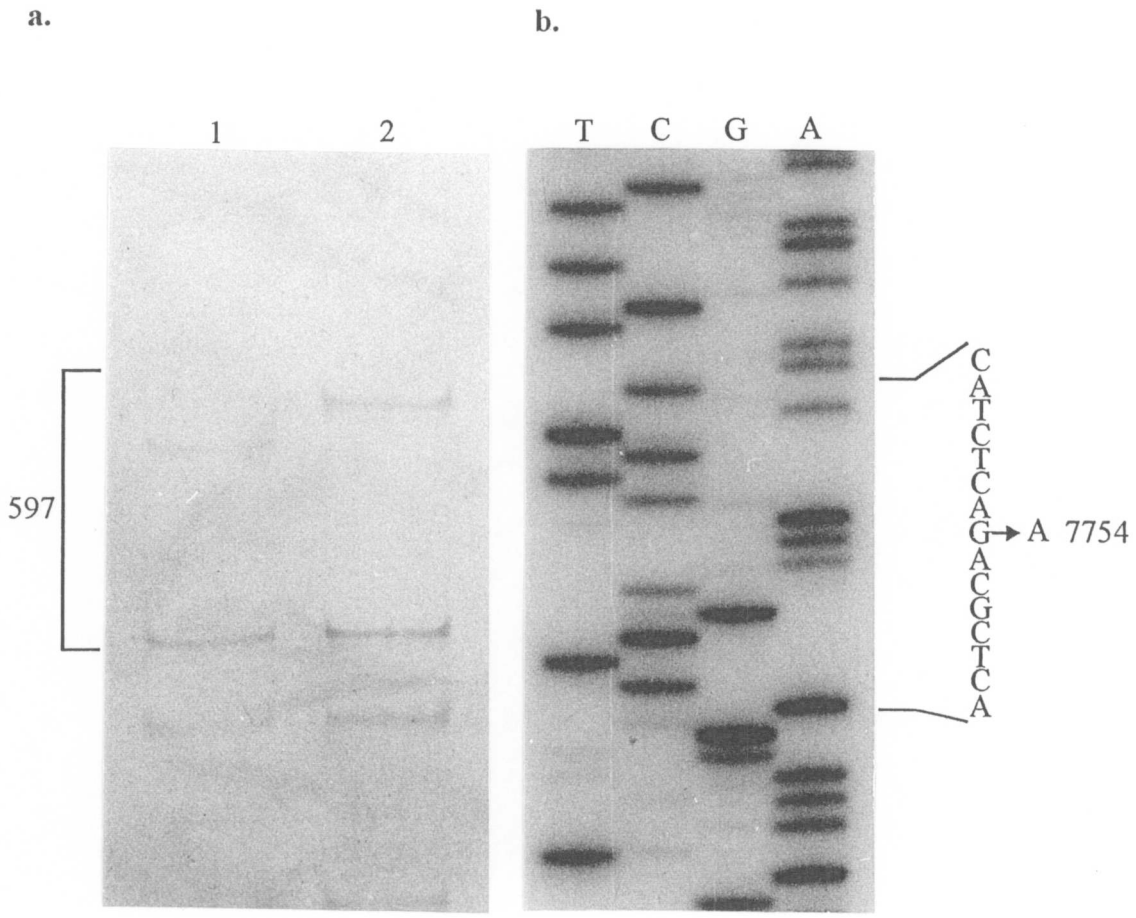


Figure 3.9 **a**. Pattern 7 from COII amplicon screening of semen sample AM470K. Lane 1 shows the two bands representing the 597 bp fragment from *Bst*N1 cleaved AM470K COII amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence). Additionally the 467 bp *Bst*N1 fragment had a mobility shift indicative of pattern 3. This was the same *Hae*III RFLP as found in pattern 3 (not shown). Restriction fragments containing mobility shifts are the same as pattern 5 (Fig. 3.7 **b**). The changes causing patterns 6 were in a 278 bp *Acc*1/*Hae*III fragment between nt 7610 and nt 7888. A 389 bp *Hae*III fragment (nt 7499 to nt 7888) was cloned containing the 278 bp *Acc*1/*Hae*III fragment. Positive clones, 165-5-1 and 165-6-7, were sequenced. **b**. Portion of sequencing gel of clone 165-5-1 showing the G → A transition at nt 7754.

In addition to 7 patterns caused by homoplasmic changes in the COII amplicon there were three patterns caused by heteroplasmic changes. Pattern 8 in semen sample AM433B is characterised by an extra band associated with the 597 bp *Bst*N1 fragment doublet as seen in Figure 3.10. This band was reproducible in four different PCRs from four different DNA samples. When AM433B COII PCR product was cut with *Acc*1, heteroplasmic bands were associated with 548 bp and 180 bp fragment bands on an SSCP gel (Fig. 3.11). When AM433B COII PCR product was cut with *Hae*III, heteroplasmic bands were associated with 300 bp and 389 bp fragment bands on an SSCP gel (Fig. 3.6b). These results indicated that at least two heteroplasmic changes were present in this semen sample. As only one extra band was associated with the 597 bp *Bst*N1 fragment it was assumed that the changes were on the same DNA molecules.

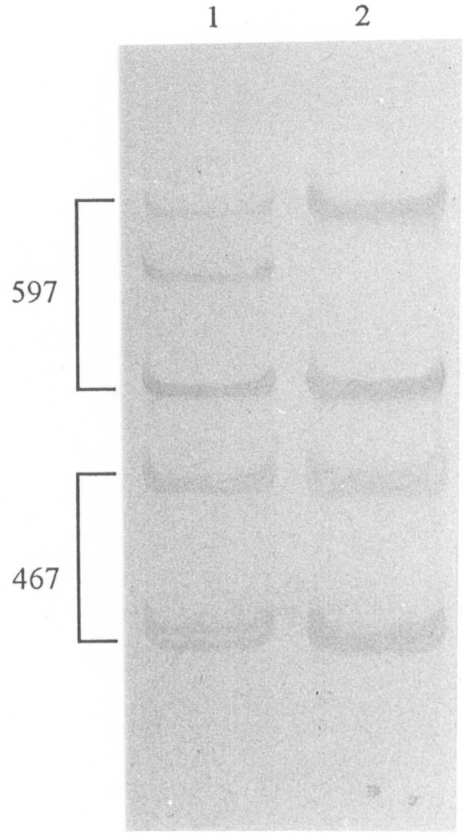


Figure 3.10. Heteroplasmic SSCP pattern of semen sample AM433B. Lane 1 shows three bands representing the 597 bp fragment from *Bst*N1 cleaved AM433B COII amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence) that has only 2 bands.

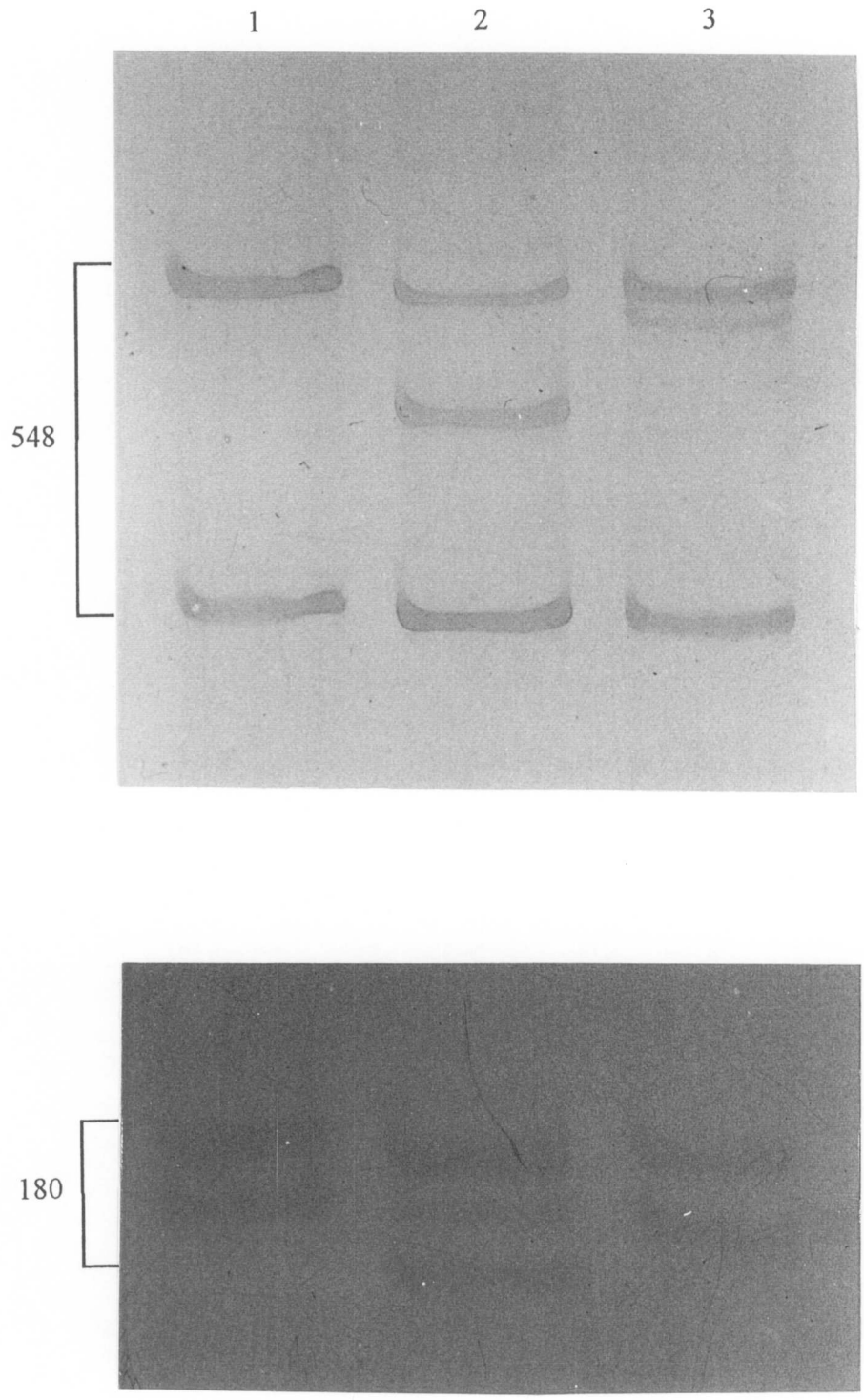


Figure 3.11. Heteroplasmic patterns of semen samples AM433B and 94-107 when cleaved with *Acc*1. Lane 1, control semen sample; Lane 2, AM433B; Lane 3, 94-107. Only the 548 nt and 180 nt bands representing the respective *Acc*1 digest fragments are shown.

The change within the 548 bp *Acc*1 fragment was assumed to be between the *Acc*1 site at nt 7610 and the *Bst*N1 site at nt 7975. The second was in the 180 bp *Acc*1 fragment between nt 7430 and nt 7610 (Fig. 3.6 b) A *Bst*N1/*Acc*1 double digest of AM433B COII PCR product was cloned in duplicate. Clones containing the 365 bp fragment between nt 7610 and nt 7975 were differentiated from other positive clones by their size (when excised from the vector with *Pvu*II) and by cutting possible positives with *Asp*1 (insert) and *Bam*H1 (vector). The inserts were excised, using *Pvu*II digestion, from 15 clones with the correct insert and all with the same orientation and were analysed by SSCP to obtain clones for sequencing. Two SSCP patterns were obtained. Sequencing of the clones with these two patterns revealed that one SSCP pattern was caused by a G → A transition at nt 7789 represented by clones 147-3-11, 39 and 147-4-1 (from two different cloning reactions). Clones with the other SSCP pattern (149-3-45 and 149-4-66) had no changes with respect to the Cambridge sequence. Clone 147-3-8 had an artefactual T → C transition at nt 7943 even though it had the same SSCP pattern as clones 149-3-45 and 149-4-66.

The inserts from clones containing the 180 bp fragment between nt 7430 and nt 7610 were excised, using *Pvu*II digestion. Ten clones with the correct insert were analysed by SSCP to obtain clones for sequencing. When sequenced, clones 147-3-1 and 147-4-16 had a C → T transition at nt 7476 whilst clones 147-3-6 and 147-4-13 had a sequence identical to the Cambridge sequence. Therefore the changes causing pattern 8 are at a minimum, a C → T transition at nt 7476 and a G → A transition at nt 7789.

Pattern 9 was characterised by the appearance of an extra band associated with the 597 nt *Bst*N1 fragment doublet of sample 94-107 (Fig. 3.12). This band was reproduced from six different PCRs from four different DNA preparations from this semen sample. To further characterise this pattern COII PCR product from 94-107 and controls was cleaved with *Acc*1 and the fragments were analysed on SSCP gels. The extra heteroplasmic band was now associated with the 548 nt *Acc*1 fragment bands (Fig. 3.11) and so the change was in a 365 bp fragment between nt 7610 and nt 7975. This fragment was cloned in duplicate. The inserts were excised, using *Pvu*II digestion, from 65 recombinant clones and analysed by SSCP to identify clones for sequencing (Fig. 3.13). Positive clones were cleaved with an *Asp*1/*Bam*H1 double digest to delineate

insert orientation. The *Asp*1/*Bam*H1 digests and the preceding SSCP's gave the patterns shown in Figure 3.13 where the bottom bands define the insert orientation and the top bands show the heteroplasmic difference between mutant and normal. Forty three (66%) of the positive clones had a normal pattern whilst 22 (34%) of the positive clones had a mutant pattern.

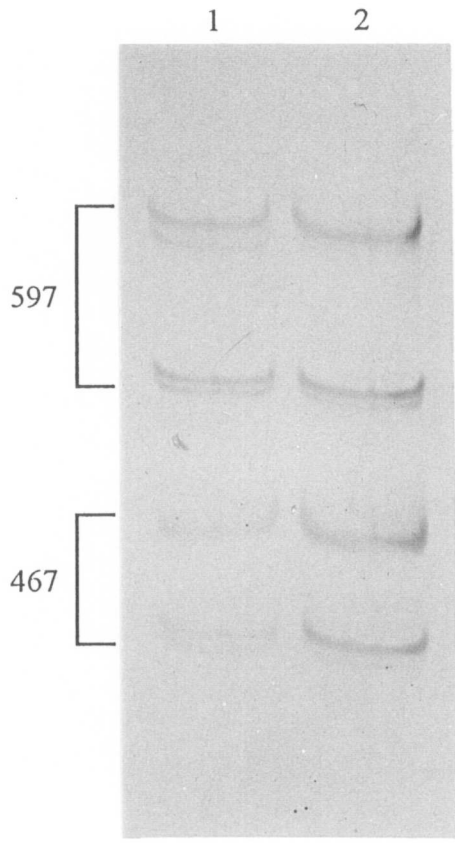
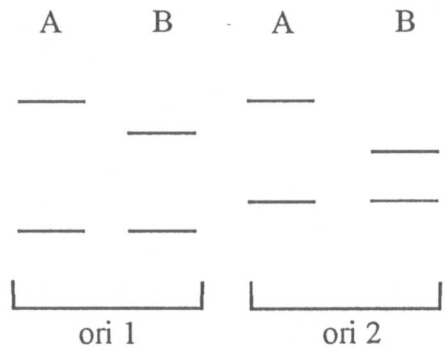


Figure 3.12. Heteroplasmic SSCP pattern of semen sample 94-107. Lane 1 shows three bands representing the 597 bp fragment from *Bst*N1 cleaved 94-107 COII amplicon as visualised on a 5% SSCP gel, compared with lane 2, a control sample with the most common pattern (assumed to have the Cambridge sequence) that has only 2 bands.

a.



b.

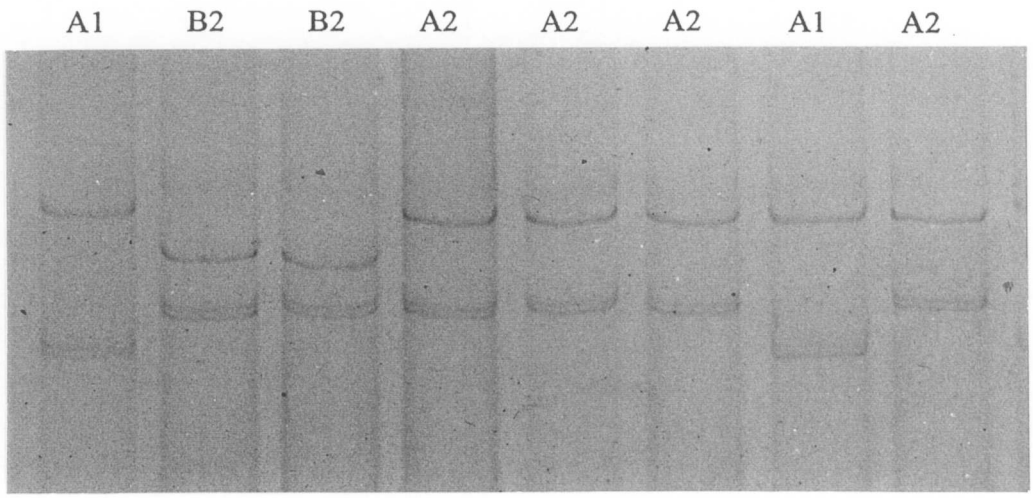


Figure 3.13. SSCP analysis of 94-107 365 bp COII amplicon clones cleaved with *Pvu*II showing: **a**. Diagrammatical representation of SSCP patterns observed. The mobility of the bottom bands define the insert orientation (ori 1 or ori 2) and the mobility of the top bands show the heteroplasmic difference between mutant and normal (A and B). **b**. a cross section of patterns represented in (**a**) as seen by SSCP analysis. Clones are classed by DNA type (A or B) and insert orientation (1 or 2). No B1 clones are shown.

A number of these clones, representing different SSCP patterns, were sequenced. The mutant sequence was represented by five clones from two PCRs showing seven changes with respect to the Cambridge sequence, with the normal being the same as the Cambridge sequence (Table 3.3).

Table 3.3 Consensus changes found in the COII mutant clones of 94-107. Syn means synonymous, or no change.

Nucleotide	Change	Amino Acid
7650	C → T	Thr → Ile
7705	T → C	syn
7757	G → A	Ala → Thr
7810	C → T	syn
7868	C → T	Leu → Phe
7891	C → T	syn
7912	G → A	syn

A number of clones had additional changes that were seen only once and so were assumed to be PCR artefact (Appendix 4). Clone 3-1-4 did not have a G → A at nt 7757 but did have other changes (although not sequenced in its entirety). Therefore, within the COII amplicon of sample 94-107 there are two populations of DNA differing by at least 7 point mutations.

Pattern 10 was initially thought to be the same as pattern 3. Sample TM showed an identical shift in the top *Bst*N1 467 bp fragment band to pattern 3, but when cleaved with *Hae*III, the pattern 3 *Hae*III RFLP was not present, and in fact no mobility differences were seen at all. When cleaved with *Acc*1, however, the 284 bp fragment bands were condensed when compared to controls.

Due to the proximity of primers HMTL817 and HMTH844, the COII amplicon of semen sample TM was sequenced directly using cycle sequencing. Note that whilst primer HMTL817 falls within the region of interest, the sequence upstream of this primer was read from the other primer. The sequence obtained from TM PCR product was ambiguous, even when repeated. The sequence between HMTH844 and nt 8277 appeared 'fuzzy' whilst sequence between HMTL817 and nt 8277 was clear (Fig. 3.14 a). This implied a length heteroplasmy centred around nt 8277 which would then make sequence down stream from this site ambiguous. Given that there are C tracts in this non

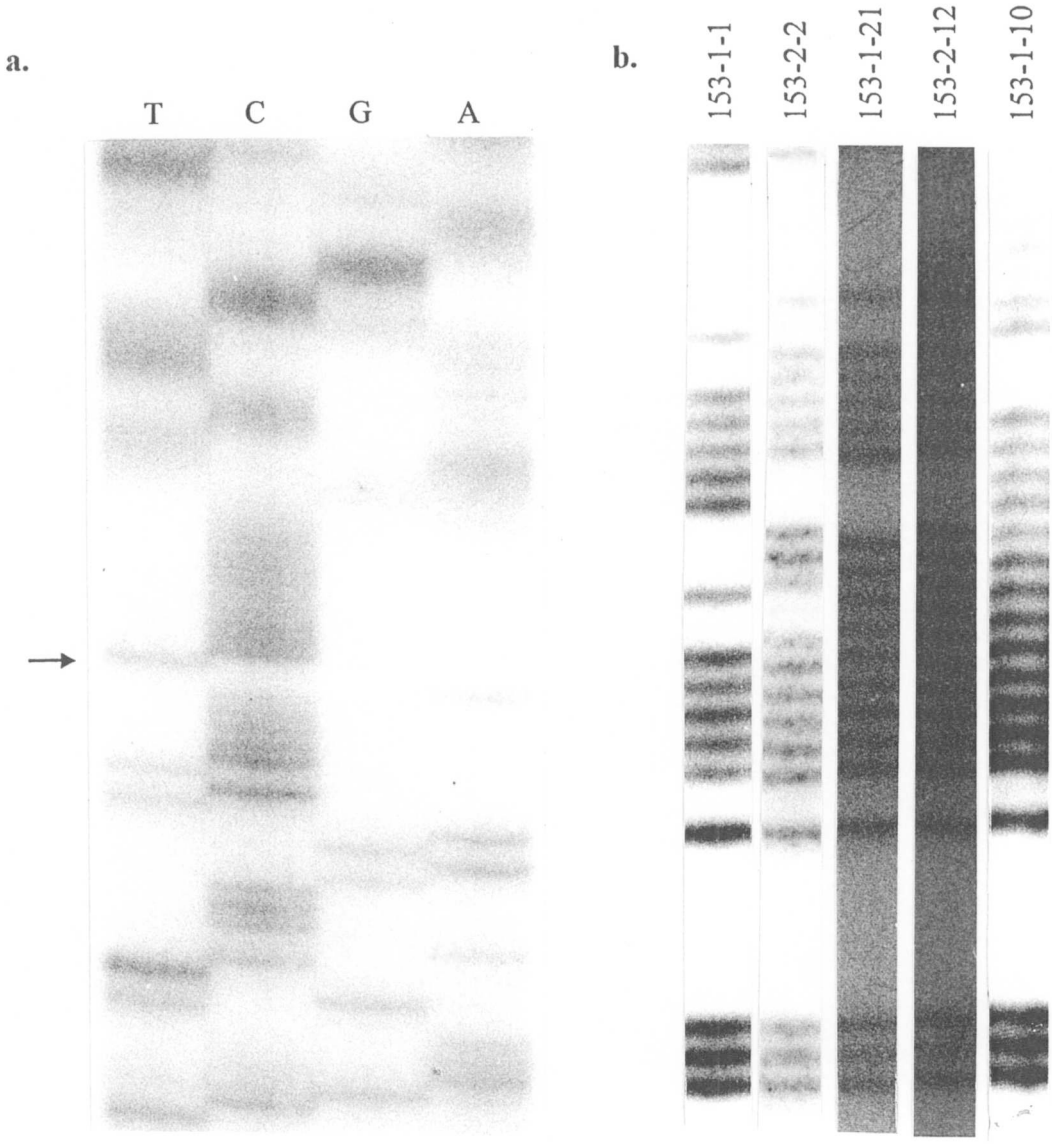


Figure 3.14.**a**. Cycle sequencing of pattern 10 from semen sample TM priming off HMTH844. Note the blurred sequence at the top, with relatively clear sequence at the bottom, indicative of homopolymeric tract length heteroplasmy. **b**. C-tracking of TM clones showing variable C_n from C_5 in lane 1 to C_{13} in lane 5.

coding area a homopolymeric C tract heteroplasmy was suspected. As stated by Marchington *et al.* (1997) in reference to mtDNA control region heteroplasmy, "Heteroplasmy of this length variation in homopolymeric tracts has not been recognised because of blurring of bands on a sequencing gel after the tract. That this was not a sequencing artefact was demonstrated by cloning and sequencing of individual clones, at which point the sequence becomes readable". Because the changes were seen between primers HMTL817 and HMTH844 these were used to amplify a short PCR product for cloning. This PCR product was cloned in duplicate as above. The inclusion of primers in the cloned sequence allowed clone inserts to be PCR amplified out of the phagmid to yield orientation independent products. Twenty nanograms of alkaline-lysis prepared phagmid DNA was used as the template for these amplifications with the PCR conditions being the same as for the initial amplification. Products were visualised on SSCP gels using standard conditions. Ten recombinant clones were sequenced (Fig. 3.15).

```
CAMBRIDGE   TTTACCCTATAGCACCCCCTC------TACCCCCTCTAGAGCCC
153-1-01    TTTACCCTATAGCACCCCCTC------TACCCCCACTAGAGCCC
153-1-10    TTTACCCTATAGCACCCCCCCCCCCCCTACCCCCTGTAGAGCCC
153-2-02    TTTACCCTATAGCACCCCCCTCCC---TACCCCCTCTAGAGCCC
153-2-07    TTTACCCTATAACACCCCCTC------TACCCCCTCTAGAGCCC
153-2-08    TTTACCCTATAACACCCCCTC------TACCCCCTCTAGAGCCC
153-1-14    TTTACCCTATAGCACCCCCCCCCC---TACCCCCTCTAGAGCCC
153-1-21    TTTACCCTATAGCACCCCCCCCC----TACCCCCTCTAGAGCCC
153-2-12    TTTACCCTATAGCACCCCCCCCCC---TACCCCCTCTAGAGCCC
153-2-13    TTTACCCTATAGCACCCCCCCCC----TACCCCCTCTAGAGCCC
```

Figure 3.15. TM clone sequences between nt 8258 and nt 8294 showing point mutation and length heteroplasmy. Dashes are present for alignment purposes only. Underlining indicates mutations relevant to pattern 10.

This sequencing revealed that two different heteroplasmic mutations were present in semen sample TM. The first was a T $\rightarrow$ C transition at nt 8277. This created a C_7 homopolymeric tract. The second mutation, which is probably wholly dependant on the presence of the first, was the non-templated addition of between C_2 and C_6 (Fig. 3.14b). The presence of variable C tracts all together created the blurring of sequence on the cycle sequencing autoradiogram after this point (Fig. 3.14a). Seven out of 20 TM clones (35%) showed a normal SSCP pattern with the other 65% having varying sizes of insertion.

Discussion

Sensitivity of PCR-SSCP for mutation detection

As mentioned earlier, SSCP when optimised, may resolve 70 to 80% of differences between fragments. As with the ATPase amplicon screening, two of the three *Bst*N1 cut COII fragments were outside the reported optimal size for SSCP mutation detection (Beier, 1993). The position of the changes found on the fragments (Fig. 3.16) are spread widely throughout the COII amplicon but generally are not close to the fragment termini (with the exception of changes close to the *Hae*III site at nt 7888). The general lack of detection of mutations at fragment termini could be because either they are not present, or that that ends of the single stranded DNA are loose, (not bound in a secondary structure) and do not significantly contribute to the overall folding. Therefore, if a large amplicon is cleaved into many smaller fragments, the ability of SSCP to detect mutations may be reduced because of the increased proportions of ends. The larger the fragment the less effect a single change has on the overall conformation leading to less chance of a change in band mobility. There is, therefore, a trade off between small optimal fragment size and the inability of SSCP to detect changes in fragment ends.

Mutation analysis

In all, seven different single base changes, a 9 bp deletion, a heteroplasmic insertion and two apparently complex base substitution heteroplasmies were characterised within the COII amplicon. Only those changes that were found in more than one sample were tested for statistically significant association to seminal parameters. The proportion of samples with pattern 2 with asthenozoospermia was found to be statistically significantly different to the sampled population proportion with asthenozoospermia.

The T $\rightarrow$ C transition at nt 8077 detected in semen sample JN has not been reported previously although the loss of an *Alu*1 restriction site across this position has

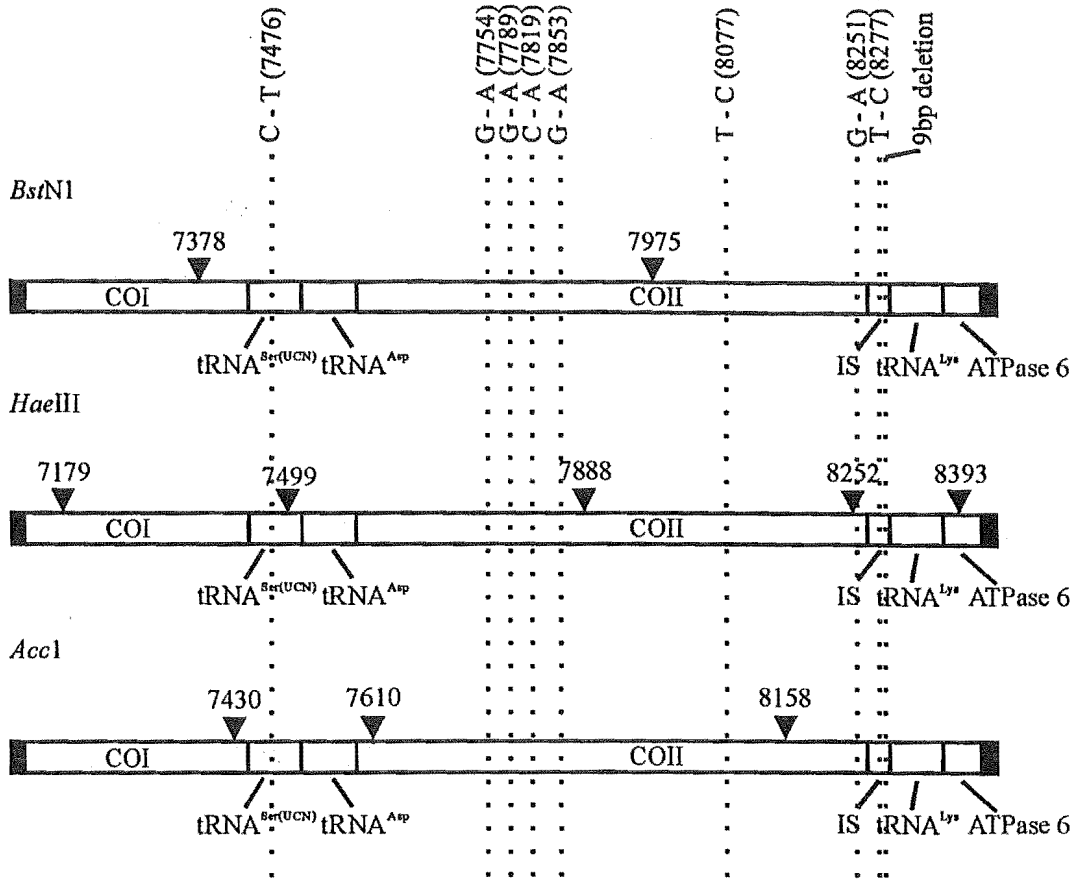


Figure 3.16. *Bst*N1, *Hae*III and *Acc*1 COII digest fragments and the position of the changes within them. Twenty bp primers are indicated in black at the ends of the COII amplicon. All genes are labelled in addition to 25 bp intergenic sequence (IS). The changes found in semen sample 94-107 have been excluded from this figure because the relative contribution of each of the changes to the overall SSCP pattern was unknown.

(individual 68; Cann *et al.*, 1987). Given that the new codon does not change amino acid assignment this is thought to be a sequence polymorphism with no pathological significance. This nucleotide position is not highly conserved, (*Pongo pygmaeus*, C (Horai *et al.*, 1992); Bovine, T (Anderson *et al.*, 1982); *Xenopus*, T (Roe *et al.*, 1985)). In the present study the frequency of this polymorphism is 0.45%.

QD788L has two changes evident within the COII amplicon. The first, a C → T transition at nt 7476 within $tRNA^{Ser(UCN)}$, has been described previously (Houshmand *et al.*, 1994; Howell *et al.*, 1995). It is unlikely that this change has pathological significance as most tRNA changes that are a primary genetic defect are heteroplasmic, whereas in QD788L this change was homoplasmic. The changed base falls in the anticodon stem of the tRNA and may actually strengthen this stem, by creating a

stronger U - A bond, theoretically creating a more stable tRNA (Fig. 3.17). Conversely, a study by Nakamura *et al.* (1995) found a pedigree with a heteroplasmic T → C transition at nt 7512 (Fig. 3.17) that weakens a strong bond in the acceptor stem. This is thought to manifest in MERFF/MELAS overlap syndrome. Whilst nt 7512 is not highly conserved, a strong bond (A - U or G - C) at this position is. The base at 7476 is neither highly conserved (Yokogawa *et al.*, 1991) nor part of a highly conserved stable bond as shown in Figure 3.17. Therefore the change at nt 7476 is thought to be a neutral population polymorphism.

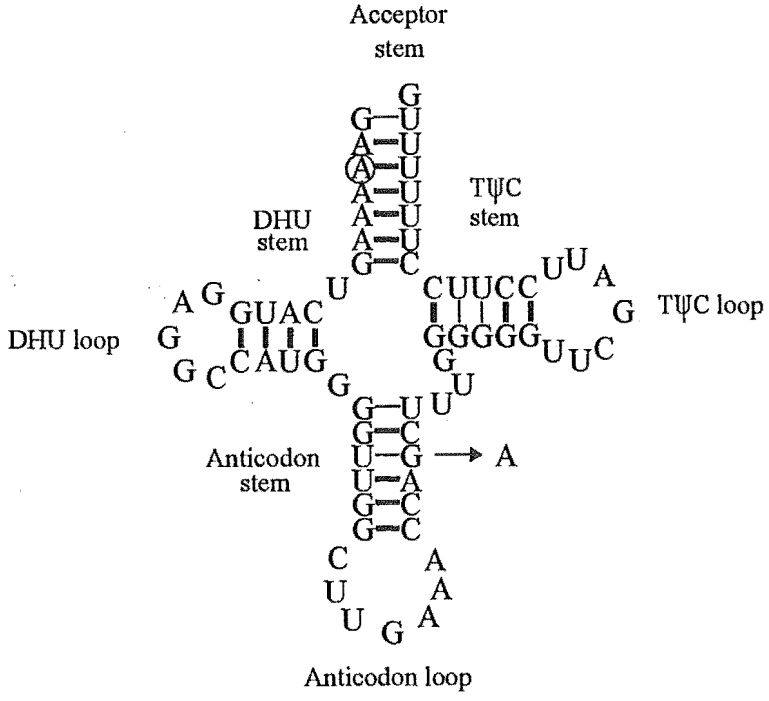


Figure 3.17. $tRNA^{Ser(UCN)}$ secondary structure (modified from Yokogawa *et al.*, 1991) showing the position of the mutation at nt 7476. (-) indicates binding affinity (the thicker the band, the stronger the bond). Note that the T → C change appears as a G → A as this is an RNA structure and $tRNA^{Ser(UCN)}$ is transcribed off the mtDNA light strand. The circled base is changed in the study of Nakamura *et al.* (1995).

QD788L has an additional G → A transition at nt 7789 that has not been reported previously. Given that the new codon does not change amino acid assignment this was thought to be a population variant with no pathological significance. *Gorilla gorilla* and *Pan troglodytes* (Ruvolo *et al.*, 1993) are both polymorphic at this position suggesting the residue is not highly conserved.

The C → A transversion at nt 7819 in semen sample SF96-279 has not been reported previously. Given that the new codon does not change amino acid assignment this was thought to be a population variant with no pathological significance. This residue is an adenine in *P. troglodytes* and a thymine in *P. paniscus* (Ruvolo *et al.*, 1993) indicating that both transitions and transversions have occurred at this site during primate evolution. In the present study the frequency of this polymorphism is 0.45%.

The G → A transition at nt 7853, seen here in semen sample CCR, has been reported previously (Chen *et al.*, 1995) in five African individuals from two different racial groups. Sample CCR has not been tested for other African haplotype markers. A G → A transition at this position will change COII amino acid Val^{90} to an isoleucine residue. Both of these amino acids have aliphatic side chains giving them a hydrophobic tendency. Adenines at this nucleotide position can be seen in *G. gorilla* and *P. pygmaeus* (Ruvolo *et al.*, 1993). Semen sample CCR remained coagulated and viscous throughout this study, whereas normal seminal samples liquefy within 20 - 30 minutes of ejaculation (Glover *et al.*, 1990). It is unlikely, however, that the G → A transition at nt 7853 in CCR had any bearing on this viscosity. As with other singly occurring mutations, in the present study the frequency of this polymorphism is 0.45%.

AM470K had two defined nucleotide changes within the COII amplicon. The G → A transition at nt 7754 has not been previously reported although a *Dde*1 site (which this position is part of) centred at nt 7750 is polymorphic (Wallace *et al.*, 1995). A G → A transition at this position will change COII amino acid 57 from aspartic acid to asparagine. Asparagine is the amide derivative of aspartic acid. Interestingly, both the guanine at nt 7754 and Asp^{57} are moderately conserved (primates, G, Asp (Ruvolo *et al.*, 1993); Bovine, G, Asp (Anderson *et al.*, 1982); *Xenopus*, G, Asp (Roe *et al.*, 1985); sea urchin, G, Glu (Cantatore *et al.*, 1989)) suggesting that these residues may be important. In beef heart COII, Asp^{57} is thought to be bonded by a hydrogen atom to the phosphate of a phosphotidyl ethanolamine just on the matrix side of the inner mitochondrial membrane (Fig. 3.2). This hydrogen bonding, however, is through the main chain imide group rather than through a side chain (Tsukihara *et al.*, 1996), thereby, theoretically allowing any amino acid to take this position. No other role has been assigned to this residue. There are examples of both pathological mutations and population

polymorphisms being caused by moderately to highly conserved aspartic acid to asparagine amino acid changes (eg. Howell *et al.*, 1993; Ozawa *et al.*, 1991). It is unlikely that this change has any phenotypic effects on semen sample AM470K's donor given the average sperm count and motility of AM470K (Appendix 1), the moderate residue conservation, the homoplasmic nature of the change and the possibility that it may have been seen as a variant before as a *Dde*1 RFLP. In the present study the frequency of this polymorphism is 0.45%.

The other AM470K change found in the COII amplicon is a G → A transition at nt 8251 which has been described and characterised in both a Japanese sample (Tanaka and Ozawa, 1994) and within a South African !Kung individual (Hsa6; Ruvolo *et al.*, 1993). Contrary to this, however, is the finding of Torroni *et al.* (1994) that the *Hae*III RFLP centred around this site (not sequenced by them) is specific to European haplotype I. The loss of the *Hae*III restriction site could be due to a mutation at any of the four nucleotides (GGCC) that make up the *Hae*III recognition site, thus leaving doubt that the G → A at nt 8251 is the primary cause of the European haplotype I *Hae*III RFLP. It has additionally been shown, however, that as well as losing a *Hae*III site, individuals with the European I haplotype gain an *Ava*II site (Torroni *et al.*, 1994) with a cleavage site at nt 8249. For this to occur the mutation must be either a G → A or G → T at nt 8251. Given that transitions are many times more frequent than transversions (Brown *et al.*, 1982), the former of the two is more likely. If this is the case then the *Hae*III RFLP is not specific for European haplotype I, but has occurred multiple times in human evolution in many ethnic groups. Supporting this idea is the study of Shoffner *et al.* (1993) that places Caucasian individuals with haplotype I on three different branches of a phylogenetic tree, suggesting three independent origins for this mutation (Shoffner *et al.*, 1993; Fig. 4 therein). None the less, this change appears to be relatively common among Europeans with reported frequencies of 7.4% (Torroni *et al.*, 1994) and 7.5% (Shoffner *et al.*, 1993). Given that this change occurs at the third position of a glycine codon within COII no amino acid change will occur. The G → A transition at nt 8251 is assumed to be a neutral population polymorphism. In the present study, 4.9% of the samples contained this change (11/223). This is not significantly different to the population percentage of

between 7.4 and 7.5% ($\chi^2 = 2.04$, df = 1, $P > 0.05$). No significant association was found between sperm count and motility and the presence or absence of this change (see results). AM470K was the only sample to have additional changes to the G → A at nt 8251. However, because it is assumed that the G → A at nt 8251 has occurred on more than one occasion in evolution it is not possible to say which mutation occurred first.

The 9 bp deletion in the COII/$tRNA^{Lys}$ intergenic spacer, originally thought to be an Asian specific marker, has now been found widely in Polynesian and Central and Southern African populations (Sykes *et al.*, 1995; Soodyall *et al.*, 1996 and references therein) and once in a Caucasian population (Barrientos *et al.*, 1995). This well characterised deletion results from the removal of one of two copies of a 9 bp direct repeat (CCCCCTCTA) found in this intergenic sequence (either nt 8272 to nt 8280 or nt 8281 to nt 8289). Haplotype analysis confirms that this mutation has re-occurred at least twice in African populations and that the Asian deletion arose independently from these events (Soodyall *et al.*, 1996). Up to 94% of Polynesian individuals in New Zealand carry this deletion (Sykes *et al.*, 1995) with no Caucasians in this country being reported to harbour the change. Seven semen samples in this study had this change. Samples BT and 95-37 were donated by the same individual. The individual who donated samples BT and 95-37 has Kallmann's syndrome. This is an X-linked (Xp22.3) disorder characterised by a gonadotropin releasing hormone deficiency with hypogonadotropic hypogonadism and delayed puberty. Whilst there are also other autosomal recessive and dominant forms it is unlikely that the 9 bp deletion has any bearing on this disease. It is likely, however, that the low sperm counts and motilities of semen samples BT and 95-37 are caused at least in part by Kallmann's syndrome. The proportion of samples with pattern 12, with asthenozoospermia was found to be significantly different to the population proportion at the 95% confidence limits. If the two samples whose low counts can be explained by Kallmann's syndrome are removed from this analysis, leaving an effective sample size of three (excluding missing data) this becomes non-significant (exact binomial probabilities, $p = 0.0526$, two tailed test with confidence limits of 95%). It is likely that the six individuals donating these samples have Polynesian maternal origins although this has not been tested further. No evidence has ever been found that this deletion has deleterious effects.

Three examples of COII amplicon heteroplasmy were seen during this screening. Sample TM has a homopolymeric C tract length heteroplasmy within the COII/tRNA^{Lys} intergenic spacer. Homopolymeric tract heteroplasmies have previously been found in the human D-loop between nt 16184 and nt 16193 (Bendall and Sykes, 1995; Marchington *et al.*, 1996; Marchington *et al.*, 1997). No homopolymeric tract heteroplasmies have been found in the COII/tRNA^{Lys} intergenic spacer, however, analogous length changes have been seen in homoplasmy (Wrischnik *et al.*, 1987; Ballinger *et al.*, 1992). Only a length change, first seen by Cann and Wilson (1983), has been fully characterised (Wrischnik *et al.*, 1987). This change (a C_{11} tract) and the ones presented here are probably caused by the same double hit mutational event. Firstly a T → C transition occurs at nt 8277. This is heteroplasmic in semen sample TM but homoplasmic in the C_{11} tract (Wrischnik *et al.*, 1987). Secondly this transition creates a C_7 tract that allows replication slippage to add C_4 (Wrischnik *et al.*, 1987) or C_n (present study). Without the transition at nt 8277 the replication slippage does not occur (Fig. 3.19)

CAMBRIDGE	TTTACCCTATAGCACCCCCTC-TACCCCCTCTAGAGCCC
FIRST CHANGE	TTTACCCTATAGCACCCCCCC-TACCCCCTGTAGAGCCC
SECOND CHANGE	TTTACCCTATAGCACCCCCCCC$_n$TACCCCCTGTAGAGCCC

Figure 3.19. Schematic representation of mutational progression in COII/tRNA^{Lys} of TM. Sequence shown is nt 8258 to nt 8295. Dashes are used for alignment purposes only. Underlining indicates T → C transition at nt 8277.

Interestingly, Cann and Wilson (1983) had difficulty sequencing this region (Gomer *et al.*, 1985) possibly suggesting that their C_{11} tract is actually also heteroplasmic, but that not enough clones were analysed by Wrischnik *et al.* (1987) to recognise this. Other groups have recognised additions in his region that may well be caused by the same mechanism but these have not been cloned or sequenced (Ballinger *et al.*, 1992). The changes appear to have occurred more than once in evolutionary history (Ballinger *et al.*, 1992) and can be fixed to homoplasmy rapidly (Wrischnik *et al.*, 1987). All three previously identified individuals with an addition in this region have an Asian mtDNA haplotype (the haplotype of TM's donor has not been identified). This homopolymeric tract heteroplasmy is not thought to have any phenotypic effect on TM's donor given that the heteroplasmy is within an intergenic spacer with no known function.

This change is probably under slight negative selection pressure as it would create a longer (slower replicating) mtDNA molecule.

The TM SSCP pattern for the COII amplicon (pattern 10) was almost identical to pattern 3. Both of these patterns were indicative of homoplasmic mutations. Sequencing revealed, however, that 35% of the TM DNA is normal (see results) and yet no normal pattern was seen. When TM COII PCR product was electrophoresed on a denaturing polyacrylamide gel, there was no indication of variable length heteroplasmy (results not shown). Post cloning, however, the variable lengths were visible on SSCP gels. Given that the PCR and cloning reactions were replicated with the same results, the heteroplasmic T → C transition at nt 8277 is thought to be a real change. Subsequent length variants could be an artefact of the cloning. Whilst this satisfies the lack of length variation on PAGE gels prior to cloning it does not satisfy the lack of a heteroplasmic pattern in the original COII screening. This point is still unresolved and requires further work to clarify. TM's donor additionally has cryptorchidism (retention of one or both testes within the abdomen) which may account for the reduced sperm count of this semen sample (Appendix 1).

Changes seen in heteroplasmy in sample AM433B (C → T at nt 7476 and G → A at nt 7789) are the same changes seen in homoplasmy in sample QD788L. The sample's donors are not obviously related (I.L Sin, pers. comm.) and it is considered highly unlikely, that the changes in AM433B arose independently of QD788L, and presumably each other, in equal proportions to each other and on the same strands of DNA. It is assumed that AM433B is a mixture of QD788L and a normal sample and that the heteroplasmy in sample AM433B is artificial. The only other viable source of semen sample AM433B heteroplasmy is through limited paternal inheritance but as the two individuals appear unrelated this is unlikely. Sample AM433B had a sperm count of 49 x 10^6 sperm/ml, a motility of 36% and a donors age at donation of 38 years. Sample AM433B was obtained from a Percoll gradient, so contained only the most morphologically normal sperm from the original semen sample.

The apparent extreme heteroplasmy in sample 94-107 was initially an enigma. Sample 94-107 has a sperm count of 349 x 10^6 sperm/ml, a motility of 66% and a donors

age at donation of 39 years. Davis *et al.* (1997), reported multiple heteroplasmy in human COI and COII genes in a group of patients with late onset Alzheimers. Of note, two of the changes found by Davis *et al.* (1997) were a C → T transition at nt 7650 and a C → T transition at nt 7868 also found in the present study. The other 94-107 COII changes were not found in the study of Davis *et al.* (1997). Whilst screening of the *Bst*N1 94-107 COII amplicon showed heteroplasmy in the 597 nt bands only, Davis *et al.* (1997) found heteroplasmy spread within both the COI and COII genes raising the possibility that the same could be true for sample 94-107. This was analysed further, the results of which are reported in Chapter 4.

Of the seven homoplasmic point mutations found in the COII screening study, three have been characterised before with another two falling in known polymorphic restriction sites. With high resolution RFLP analysis it is unlikely the other two sites (at nt 7789 and nt 7819) would have been detected as they do not fall within the recognition sequences of the commonly used restriction enzymes (*Alu*1, *Ava*II, *Dde*1, *Fnu*DII, *Hae*III, *Hha*1, *Hin*f1, *Hpa*1, *Hpa*II, *Mbo*1, *Rsa*1 and *Taq*1). The nt 7819 change can be detected by the loss of an *Mnl*1 restriction site (normally cleaving at nt 7827) but the nt 7789 change does not alter or create a known restriction site and so cannot be detected by RFLP analysis.

Chapter 4

ANALYSIS OF SAMPLE 94-107 APPARENT HETEROPLASMY

Introduction

Heteroplasmy prevalence and maintenance

In 1985 Monnat and Loeb isolated, cloned and sequenced numerous copies of the human mtDNA COIII gene from five normal humans. Over 49 kilobases from 248 clones were sequenced to find a single base difference within one individual. Later Laipis *et al.* (1988) and Koehler *et al.* (1991) demonstrated that in Holstein cows, mother-daughter pairs can be homoplasmic for different bases at the same position suggesting a segregation of heteroplasmy to homoplasmy within one generation. These studies suggest that heteroplasmy is rare and rapidly segregated. Also, as heteroplasmy is most often found when screening for disease causing mutations, it has predominantly been associated with mtDNA diseases. As mentioned previously, there are two types of heteroplasmy associated with mtDNA. Base substitution heteroplasmies have traditionally been single nucleotide substitutions, where maternally inherited, caused by replication mismatch or reactive oxygen species damage during oogenesis, or by nuclear factors where inheritance is Mendelian.

Length heteroplasmy, as seen in the D-loop (Bendall and Sykes, 1995) and now in the COII - $tRNA^{Lys}$ intergenic spacer (Chapter 3), is not normally disease associated. This heteroplasmy is presumably created by replication slippage of homopolymeric tracts, and whilst in non-coding regions is probably selectively only slightly negative due to the molecule's increased size. Replicative segregation of length heteroplasmies has been the main tool in ascribing a bottleneck mechanism to heteroplasmy inheritance. The bottleneck (purifying selection), theoretically acting on primary oogonia, is proposed to allow for the selective amplification of different types of mtDNA (reviewed by Poulton, 1995). This could account for phenomena such as the rapid segregation of mtDNA in Holstein cattle. Conversely, others have demonstrated, using model fusion systems in mice, that all segregation phenomena can be explained by random genetic drift (eg. Jenuth *et al.*, 1996). However, as Lightowlers *et al.* (1997) state "Whether segregation

really does occur by random genetic drift or through a complex 'sampling' process not yet understood, all models predict rapid segregation".

Whilst rapid segregation may be the norm, not all heteroplasmies follow this course. Howell *et al.* (1992) found slow mitochondrial DNA segregation in a silent heteroplasmic change in ND6 (G → A at nt 14560) that was found by chance when looking at a LHON pedigree. Interestingly, this site is polymorphic in humans suggesting there is no real selective advantage to maintain heteroplasmy at this silent site. In 1996, Howell *et al.* described another silent heteroplasmic polymorphism, seen within a large maternal lineage, at nt 14470. This was fixed to homoplasmy in one branch of the family, whilst within another branch it remained heteroplasmic or was lost (Howell *et al.*, 1996; Fig. 2 therein). Vilkki *et al.* (1990), like Howell and coworkers (1996), showed that a change within a large maternal lineage could be fixed rapidly, or remain heteroplasmic. Whilst the mutation described by Howell and coworkers (1996) is presumably silent, the mutation described by Vilkki and coworkers (1990) at nt 11778 is a primary LHON mutation. The same phenomenon can be seen in artificially created heteroplasmic mice strains, where heteroplasmy can be maintained for a number of generations (Meirelles and Smith, 1997). Thus silent and deleterious mutations can both be fixed rapidly, or retained in heteroplasmy. This observation suggests either an inefficient bottleneck or that the segregating unit is a heteroplasmic mitochondrion as opposed to a singleton mtDNA. This phenomenon can also be explained by the continued creation of heteroplasmy due to the presence of a nuclear propensity factor (e.g. Suomalainen *et al.*, 1995).

Multiple site heteroplasmy

Multiple site heteroplasmy can be divided into three classes, all of which are inter-related. The first of these, where mutations fall on different strands leading to triplasmy (multi-plasmy), has been demonstrated previously. Comas *et al.* (1995) found two heteroplasmies in the human D-loop within one individual, but these appeared to be in different proportions and were never shown to be on the same strand of DNA. Howell *et al.* (1996) demonstrated an inherited form of triplasmy in a large maternal lineage, showing that the mutations never fell on the same strand.

The second class, and probably the most common, is likely an extension of the first. Where further mutations occur in triplasmic individuals or lineages, some will fall on DNA strands already containing mutations. This leads to a state of multi-plasmy with mutational progression as has been observed in bat sequence variation (Petri *et al.*, 1996). Jazin *et al.* (1996) showed that heteroplasmy was a natural state in human brains (non-dividing with a high energy metabolism) and that the level increased with age with multi-plasmy and mutational progression common.

Probably the rarest, and most unconventional, class of multiple site heteroplasmy is when all of the mutations fall on the same DNA strands. This effectively gives two increasingly different populations of mtDNA within an individual, with or without intermediates. Bendall *et al.* (1996), whilst studying mono and di-zygotic twins, found one individual who was slightly heteroplasmic for two different bases and because, within clones, they were never found apart, it was assumed that they were on the same strands. No intermediates were found, although the number of clones analysed was small. Semen sample TM (Chapter 3) is an example of this type of multiple change heteroplasmy although the presence of one change is directly dependent on the other. Magoulas and Kouros (1993) showed that a proportion of Mediterranean anchovies (*Engaulis encrasicolus*) had two distinct populations of mtDNA differing by at least 7 restriction sites. The authors validly attributed the difference to bi-parental inheritance as they found individuals in the population of Mediterranean anchovies that were homoplasmic for one mtDNA type or the other (and in fact the two types of homoplasmy were common). Very recently Davis *et al.* (1997) demonstrated a seemingly common six base heteroplasmy within human COI and COII genes that they associated with late onset Alzheimers. All six mutations were frequently on the same DNA strands, with the only variation being occasional intermediates, with DNA strands (represented by clones) missing one or more of the changes. In a note added in proof they suggest there are similarities between the mutant strands they found and the wild type mtDNA sequences of the great apes *Pan troglodytes*, *Gorilla gorilla* and *Pongo pygmaeus* and state that "this human mutant DNA molecule, therefore, may represent a transitional mtDNA of ancient origin". Hirano *et al.* (1997) and Wallace *et al.* (1997), when analysing the data of Davis *et al.* (1997), found the sequence divergence and the fact that the so called 'mutant' molecules have phenotypic effects even though they are found in closely related

species, disconcerting. Both groups analysed mitochondrial free cells and found that a 'mutant' sequence containing the changes described by Davis *et al.* (1997) could be amplified, suggesting that this 'mutant' sequence is a nuclear pseudogene. Both groups demonstrated that within the bounds of the COI and COII genes, this pseudogene contains a large number of sequence variants, including all but one of the changes found within the 94-107 COII 'heteroplasmy' in the present study (Chapter 3; pattern 9). That Davis *et al.* (1997) found only 6 of the 38 changes within the COI and COII pseudogenes reported by Wallace *et al.* (1997) suggests that either Davis and coworkers observed different pseudogenes or that they overlooked the additional changes. Neither Davis *et al.* (1997), Hirano *et al.* (1997) nor Wallace *et al.* (1997) extended their studies to find the extent of the nuclear insertion beyond the COI and COII gene boundaries.

The aim of the present study was to extend the analysis of the 'heteroplasmy' seen within the COII amplicon from semen sample 94-107 with the aim of confirming the pseudogene status of this mutant DNA. Conventional means of confirming pseudogene status including genomic DNA Southern hybridisation (Zischler *et al.*, 1995) and mitochondrial-free cell DNA amplification (Hirano *et al.*, 1997; Wallace *et al.*, 1997) were not possible because of a shortage of semen sample. Other less robust means were therefore used in this analysis.

Materials and Methods

To confirm pseudogene status for the mutant changes seen in COII amplicon of sample 94-107 two different approaches were taken. The first involved differential cell extraction, and the second, further genomic amplification.

Sperm nuclei are surrounded by highly thiolated (ramified) proteins, meaning that chromosomal DNA can generally only be purified if the extraction mixture contains DTT. In this way differential extraction can be carried out on sperm by prior extraction with non-DTT containing buffers, pelleting the sperm nuclei and then extracting the chromosomal DNA from sperm nuclei with DTT containing buffers (Gill *et al.*, 1985). In normal seminal samples this method is not expected to give an additional fraction of purified mtDNA as the contaminating seminal white blood cells will give up their

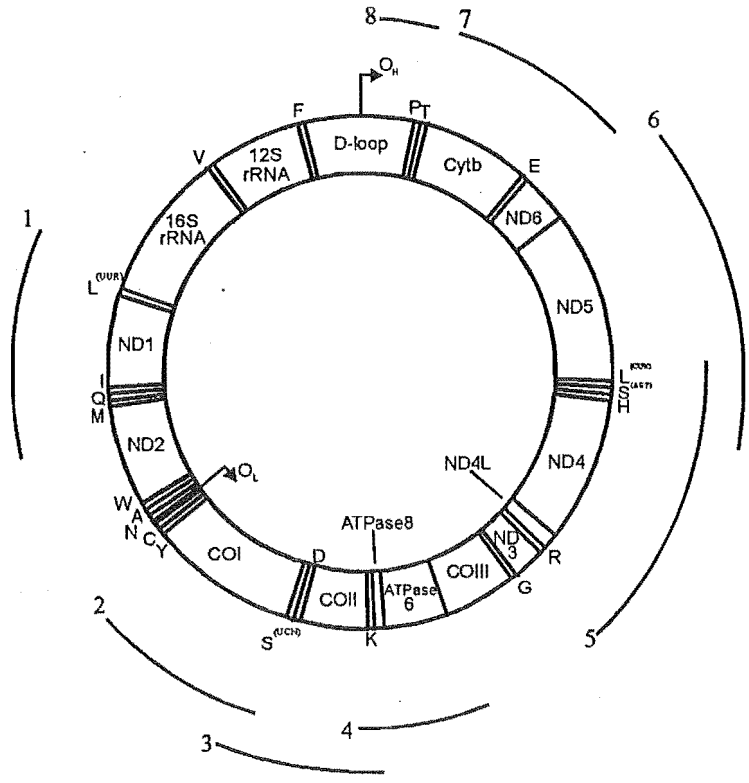
chromosomal DNA in the non-DTT extraction. Twenty microlitres of seminal fluid was centrifuged at 4°C for 10 min at 4,500 x g. The pellet was resuspended in 250 µl of extraction buffer (10 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 mM NaCl, 2% (w/v) SDS, 20 µg/ml proteinase K (Boehringer Mannheim)) and incubated at 37°C for 30 min. Sperm nuclei were pelleted by centrifuging for 5 min at 15,000 x g at room temperature. The nuclear pellet was washed once with extraction buffer and re-centrifuged as before. Nuclei were lysed in the extraction buffer with the addition of 3.9 mM DTT at 37°C overnight. DNA was purified from both the before and after DTT extraction fractions by a single phenol/chloroform extraction before being precipitated with 0.1 vol 3 M sodium acetate and 2.5 vol absolute ethanol at -20°C overnight. The DNA was centrifuged to pellet, washed with 70% ethanol, air dried and re-suspended in 20 µl of ddH_2O. The COII amplicon from semen sample 94-107 and controls was then amplified using primers HMTL712 and HMTH844 as described in Chapter 3. PCR products were cleaved with *Bst*N1 and electrophoresed on 5% SSCP gels using standard conditions.

In addition to differential extraction, other regions of mtDNA were analysed to see if the differences between the normal and 'mutant' sequence were confined to the COII amplicon (Table 4.1). PCR products from semen sample 94-107 and control samples were cleaved to completion by restriction enzymes that gave products suitable for SSCP analysis. SSCP analysis was performed using standard conditions.

Based on the SSCP analysis of amplicons presented in Table 4.1, cloning of two other amplicons was carried out for replicate PCR products from semen sample 94-107. *Alu*1 cut HMTL1161-HMTH1357 94-107 PCR product was blunt ended (Klenow end filling) and the bands of interest were gel extracted using a BioRad prep-a-gene kit. These bands were then cloned into *Sma*1 cut pBSM13+. Positive colonies were selected and sequenced as described in Chapter 3. The possibility of heteroplasmy in the HV1 PCR product created by primers HMTL1598 and HMTH1642 lead to the cloning of this amplicon. 94-107 HV1 PCR product was blunt ended (Klenow end filling) and cloned into *Sma*1 cut pBSM13+ as previously described. To select clones for sequencing, 20 ng of boiling preparation from positive clones was PCR amplified with the same conditions and primers used to create the HV1 insert for cloning (Appendix 5). This gave

Table 4.1. Amplicons used in the analysis of multiple difference 'heteroplasmy' with a diagrammatical representation of the amplicon coverage of mtDNA. Note that PCR conditions optimised for these amplicons are listed in Appendix 4.

	Primers	Nucleotides	Size	Genes included
1	HMTL321 HMTH455	3216-4557	618 bp	16s rRNA, $tRNA^{Leu(UUR)}$, ND1, $tRNA^{Ile}$, $tRNA^{Gln}$, $tRNA^{f\text{-}Met}$, 87 bp ND2
2	HMTL604 HMTH735	6040-7356	1316 bp	1316 bp COI
3	HMTL712 HMTH844	7126-8442	1316 bp	319 bp COI, $tRNA^{Ser(UCN)}$, $tRNA^{Asp}$, COII, 25 bp intergenic spacer, $tRNA^{Lys}$, 76 bp ATPase 8
4	HMTL817 HMTH934	8186-9341	1155 bp	101 bp COII, $tRNA^{Lys}$, ATPase 8, ATPase 6, 134 bp COIII
5	HMTL1012 HMTH1246	10123-12461	2338 bp	281 bp ND3, $tRNA^{Arg}$, ND4L, ND4, $tRNA^{His}$, $tRNA^{Ser(AGY)}$, $tRNA^{Leu(CUN)}$, 125 bp ND5
6	HMTL1161 HMTH1357	11616-13574	1958 bp	520 bp ND4, $tRNA^{His}$, $tRNA^{Ser(AGY)}$, $tRNA^{Leu(CUN)}$, 1214 bp ND5
7	HMTL1445 HMTH1585	14459-15858	1399 bp	214 bp ND6, $tRNA^{Glu}$, 1111 bp Cyt b
8	HMTL1598 HMTH1642	15981-16420	439 bp	43 bp $tRNA^{Pro}$, 397 bp D-loop (HV1)



orientation independent (no vector DNA) positive clones for analysis of SSCP gels as described above. Sequencing of these clones was performed as previously described.

Results

Differential extraction

Differential extraction of semen sample 94-107, subsequent COII amplicon PCR, *Bst*N1 digestion and then SSCP analysis, was performed in triplicate. The nuclear fraction SSCP pattern was similar to pattern 9 (Chapter 3) in that both mutant and normal COII products were present (Fig. 4.1) The presence of a normal COII SSCP pattern within the nuclear fraction indicated that there was still normal mtDNA associated with this fraction. The mitochondrial fraction, however, showed only the normal COII SSCP pattern. This indicated that the mutant COII pattern seen in semen sample 94-107 was restricted to the nuclear fraction and therefore was the product of a nuclear gene amplification.

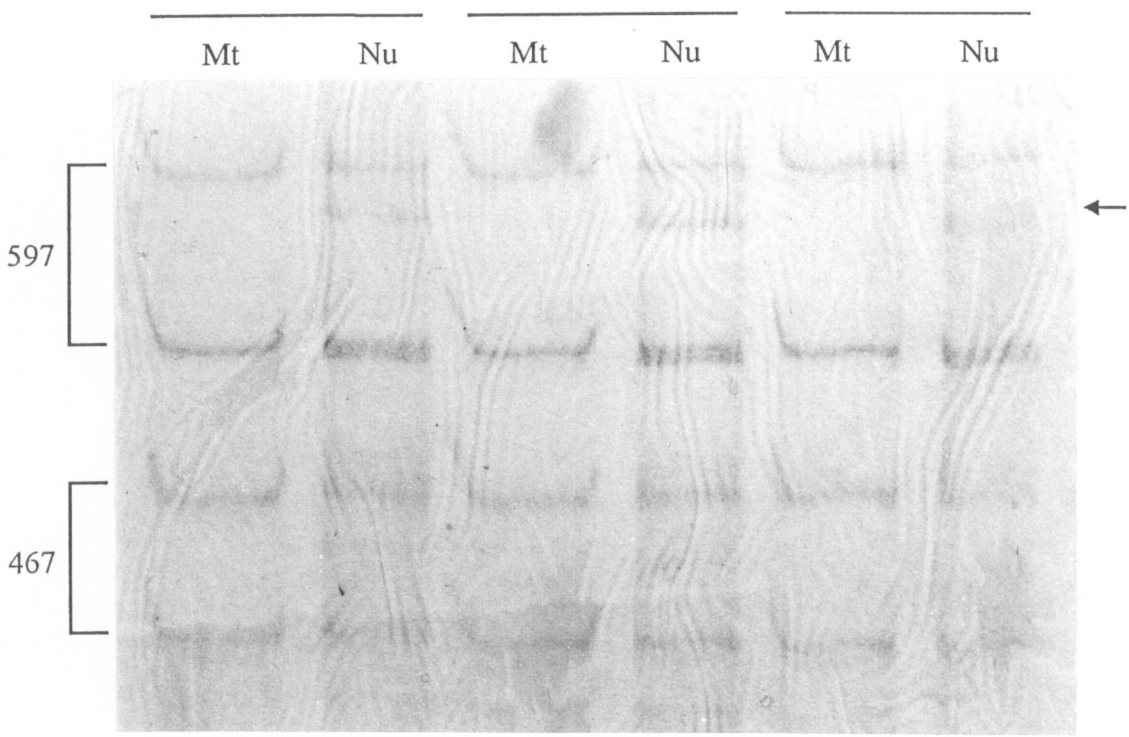


Figure 4.1. SSCP of sample 94-107 COII amplicon PCR after differential extraction. Nu is the nuclear fraction, whilst Mt is the mitochondrial fraction. The 'mutant' COII band is arrowed. Results are shown as three replicated extractions.

To determine if the heteroplasmy was confined to COII, 9931 bp of non-continuous mtDNA was screened by amplifying and cleaving specific amplicons (Table 4.1), and analysing by SSCP as before. In addition to the variations seen in the COII amplicon, only amplicons created by primers HMTL1161 - HMTH1357 (ND4/ND5) and HMTL1598 - HMTH1642 (HV1) showed heteroplasmic variation when compared to control samples.

ND4/ND5 heteroplasmy

Primers HMTL1161 and HMTH1357, spanning the ND4/ND5 amplicon, give a 1958 bp PCR product. When 94-107 ND4/ND5 product was cut with *Alu*1 and electrophoresed on a 3% agarose gel, a large number of extra fragments were visible when compared with controls. In addition to the normal *Alu*1 fragments (Fig. 4.2) of 1000 bp, 388 bp, and 278 bp, semen sample 94-107 had fragments of 890 bp, 601 bp, 531 bp, 359 bp, and 200 bp (Fig. 4.3). Control samples had an additional 143 bp *Alu*1 fragment that was missing from semen sample 94-107.

The extra *Alu*1 bands were associated with all of the ND4/ND5 amplifications from semen sample 94-107. Additionally, there was no undigested PCR product which might suggest that the extra fragments were not a result of partial digestion. Nonspecific enzyme activity was unlikely because the same banding pattern was reproducible and was not seen with any of the other semen samples tested. When the same region was digested with *Hae*III (Fig. 4.3) the expected fragments of 686 bp, 679 bp and 521 bp were still present with two additional fragments slightly larger than 686 bp and a fragment approximately 350 bp respectively.

Two rounds of cloning and sequencing were performed to identify changes to the ND4/ND5 amplicon and to obtain a genetic map of the *Alu*1 digest fragments (Fig. 4.4). In the first round the 890 bp, 601 bp and 531 bp mutant fragments were cloned and sequenced. Later cloning of the mutant 359 bp fragment and the normal 388 bp fragment confirmed changes from the above sequencing and showed that at least for some of the ND4/ND5 amplicon the normal sequence is the same as the Cambridge sequence (see Appendix 6 for clone sequences and Table 4.2 for a list of the most conserved changes).

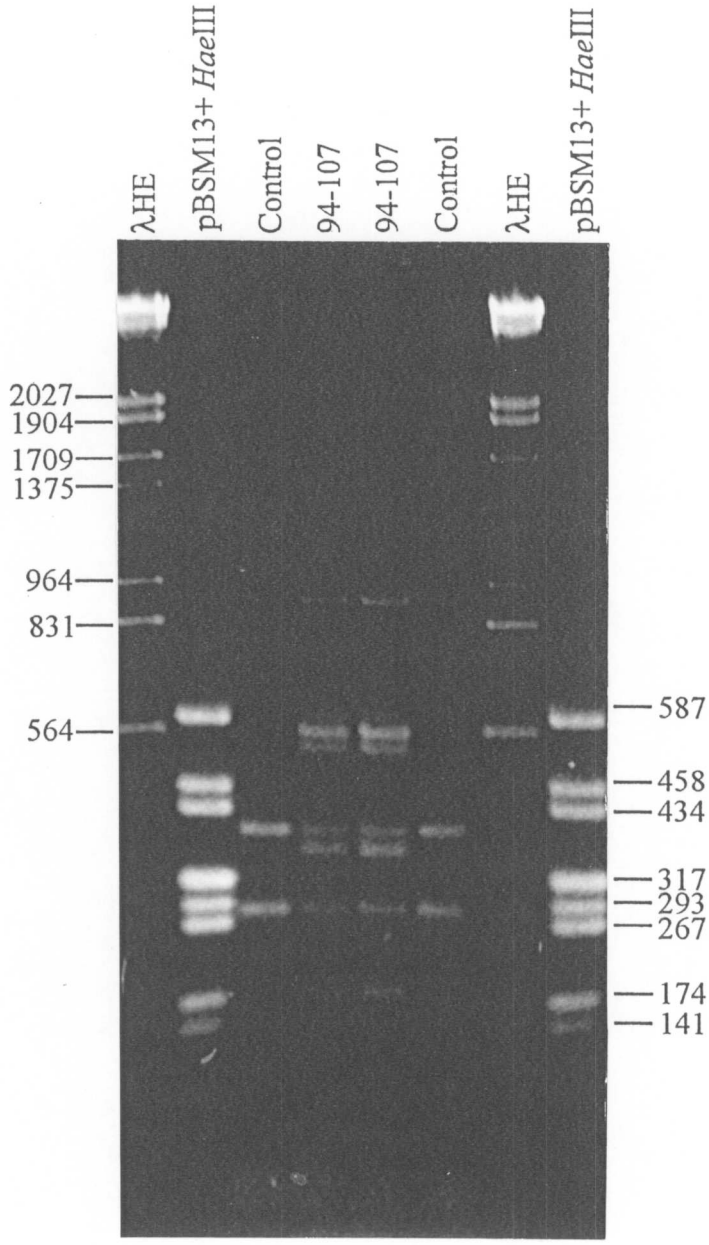


Figure 4.2. 3% agarose gel of *Alu*1 cut ND4/ND5 amplicon from 94-107 compared with control samples showing extra bands associated with 94-107 and the loss of the 143 bp *Alu*1 fragment. Markers are lambda DNA cleaved with *Hind*III and *Eco*R1 and pBSM13+ cleaved with *Hae*III.

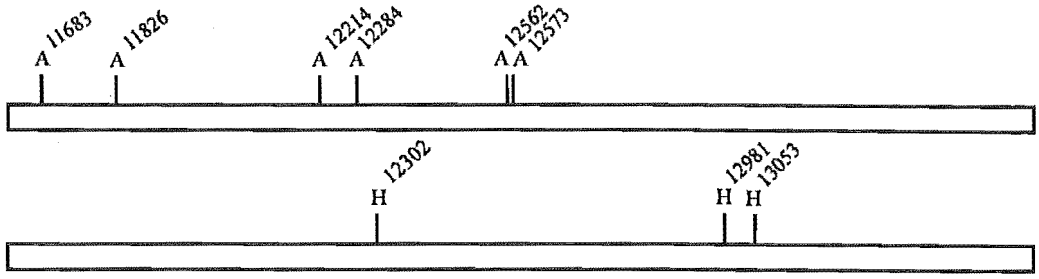


Figure 4.3 Schematic representation of restriction endonuclease cleavage sites within the ND4/ND5 amplicon for enzymes *Alu*1 (A) and *Hae*III (H). Actual cutting sites are marked.

Note that the normal 143 bp *Alu*1 fragment is not present on the agarose gel in Figure 4.2. This indicates that whilst the mutant DNA always cuts at nt 11683, the normal does not cut at nt 11683 (given that the normal cuts at nt 11826 to yield the 388 bp fragment). Thus the normal is also different from the Cambridge sequence and has a change that the mutant does not have. No polymorphisms have been reported within this *Alu*1 site previously.

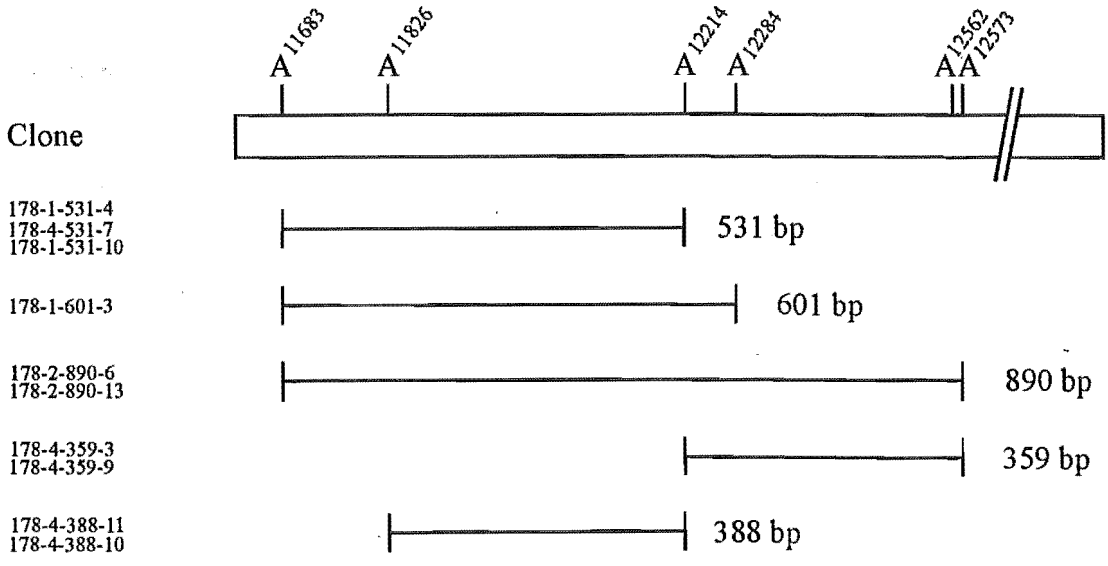


Figure 4.4. Position of sequenced clones on a fragment map of *Alu*1 cut 94-107 ND4/ND5 amplicon. Actual cutting sites of *Alu*1 sites (A) are indicated. Restriction sites at nt 12214 and nt 12284 are cut only some of the time, indicating that these sites have within heteroplasmy variation. Nucleotide 11826 and nt 12562 are never cut within mutant clones indicating that these sites always have this heteroplasmic change.

Of additional interest is the number of clones that appear to contain recombinations of mutant and normal DNA. Such recombinations are probably artificially created by *Taq* polymerase strand jumping (Bradley and Hillis, 1997). Often in these recombinations there are *Alu*1 restriction sites within the normal portion of the clone that are not cut. The reason for this is not known.

Ten clones were sequenced so that, over the 890 bp between nt 11683 and nt 12573, a consensus 'mutant' sequence was obtained. In addition to a new ND5 premature termination codon created by a T → A change at nt 12441 there were 47 other consensus point mutations (Table 4.2). In addition to the most conserved changes listed in Table 4.2, there were a number of changes that were only found in one or two clones. These changes are probably artefacts of the PCR and as such have not been used in any further analysis.

Table 4.2 Consensus changes found in the ND4/ND5 'mutant' clones of 94-107. Syn means synonymous, or no change. Dashes indicate no amino acids coded because the nucleotide change falls within a tRNA gene. Ter refers to a newly created termination codon.

Nucleotide	Change	Amino acid	Nucleotide	Change	Amino acid
11708	A → G	Ile → Val	12285	T → C	-
11719	G → A	syn	12290	A → G	-
11767	C → T	syn	12346	C → T	His → Tyr
11770	T → C	syn	12349	A → G	Thr → Ala
11788	C → T	syn	12358	A → C	Thr → Pro
11809	T → C	syn	12367	A → G	Thr → Ala
11827	T → C	syn	12372	G → A	syn
11852	G → A	Ala → Thr	12379	C → T	syn
11857	C → T	syn	12390	C → T	syn
11887	G → A	syn	12406	G → A	Val → Ile
11914	G → A	syn	12417	C → T	syn
11963	G → A	Val → Ile	12432	C → T	syn
12007	G → A	syn	12441	T → A	Tyr → ter
12013	A → G	syn	12454	G → A	Val → Thr
12018	C → G	Thr → Ser	12466	T → C	Phe → Leu
12064	C → T	syn	12469	A → G	Ile → Val
12088	C → T	syn	12474	C → T	syn
12091	T → C	syn	12501	G → A	syn
12115	C → T	syn	12503	G → A	Cys → Tyr
12136	T → C	syn	12519	T → C	syn
12189	T → C	-	12528	G → A	syn
12218	C → T	-	12540	A → G	syn
12236	G → A	-	12543	C → A	syn
12237	C → T	-	12561	G → A	syn

HV1 heteroplasmy

The 94-107 HV1 amplicon (HMTL1598 - HMTH1642) when electrophoresed, uncut and single stranded, on an SSCP gel had a heteroplasmic pattern when compared to control samples. The whole HV1 amplicon was blunt end cloned into pBSM13+ phagmid. The insert from 30 recombinant clones was amplified as described in the Materials and Methods, and then analysed by SSCP (Fig. 4.5). Two major SSCP patterns were observed and representatives of these were sequenced from duplicate cloning reactions. The normal sample 94-107 HV1 sequence is the same as the Cambridge sequence apart from a C → T change at nt 16354, which identifies this individual as possibly having a North-Eastern European origin (Table 4.3) when compared with HV1 sequences listed in the HV concordance (a database of HV sequences and racial origin) (Miller *et al.*, 1996).

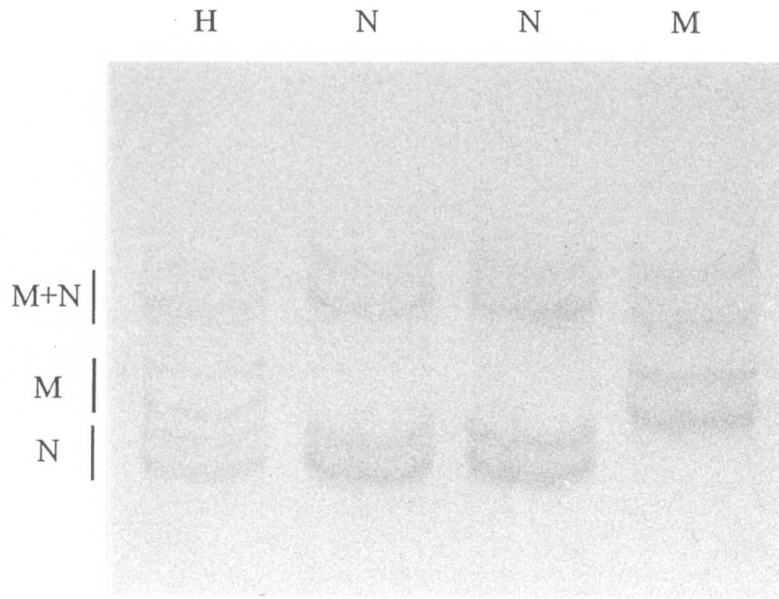


Figure 4.5. SSCP analysis of 94-107 PCR amplified HV1 clones. H represents the original heteroplasmic pattern, N, a normal pattern and M, a 'mutant' pattern.

Table 4.3. Individuals sharing the same sequence variant as the normal sequence of 94-107. Data were derived from The Concordance of Nucleotide Substitutions in Human mtDNA Control Region. Department of Anthropology, University of California at Santa Barbara (http://sscf.ucsb.edu/~kmiller) on 10.8.97 at 12pm PST. Note that whilst one of these individuals is Mongolian, this race is presumed to have much sequence variation because of its historically wide ranging population assimilation.

Individual	Population	Reference
17	Bulgar	Calafell *et al.*, 1996
38	Bulgar	Calafell *et al.*, 1996
F193	Finnish	Sajantila *et al.*, 1995
aukar5	Karelian	Sajantila *et al.*, 1995
aukar10	Karelian	Sajantila *et al.*, 1995
aukar11	Karelian	Sajantila *et al.*, 1995
17.31	Mongolian	Kolman *et al.*, 1996
SW27	Swiss	Pult *et al.*, 1994

Clones with the other SSCP pattern have a very different sequence to normal as shown in Figure 4.6. This mutant sequence has 106 transition/transversion mutations, 3 single base deletions, 1 double base deletion and 6 single base insertions compared to both the Cambridge sequence and the 94-107 normal HV1 sequence when aligned by the EMBL Homology Alignment Program in DNASIS. Thus, over the 376 bp HV1 region (Fig. 4.6) the mutant sequence shows 72% sequence identity to the normal (not including insertions or deletions). See Appendix 7 for more details.

Discussion

Differential extraction

The differential extraction technique was devised by Gill *et al.* (1985) for the separation of sperm nuclei from other cell types and debris in vaginal swabs. In this study differential extraction has been successfully used to isolate enriched fractions of nuclear and mitochondrial DNA. Subsequent PCR and SSCP analysis showed that the 'mutant' COII band was restricted to the nuclear fraction consistent with the findings of Hirano *et al.* (1997) and Wallace *et al.* (1997).

```
Cambridge       gttctttcatggggaagcagatttgggtaccacccaagtattgactcacccatcaacaaccgctatgtatt
94-107 normal   ...............................................................
94-107 mutant   ..at.c.......g...cat.........a.........c...c.....ca.c.t..-ttt...ak..a
Neanderthal     .............g.....................................g.............c
Nu insertion                                                                  .....
P. troglodytes  .................a....a.........t.....c..g....tt...t.-.............
P. paniscus     .................a....aa..g.............g....tt..ct.-t.............
G. gorilla      .............ga..a.....................g..a........-t..ttatc......g
P. abelii       .............g.c..........g......c....c....c..tttc.a.ctgg.--.........
H. lar          ..at.c.....t......cat.........a....c....c.a..c...ttc..c......-.........

Cambridge       tcgtacattactgccagccaccatgaatattgtacggtaccataaatacttgaccacctgtagtacata-a
94-107 normal   ...............................................................-.
94-107 mutant   ....ga...t....t.............a..ta....t....t.g...a..tgtaca.........tc
Neanderthal     ............tt..tt..............a.........t........t.....c........-.
Nu insertion    ...................................a.................................-.
P. troglodytes  ............................c....a........-tc..cca..t....a.........-.
P. paniscus     ..............................ac.ta....t....tc.t..a.......a..ac.....-.
G. gorilla      ....g....c.......a...........a..c..a.c....c....gtcc..t.........ac....---
P. abelii       .........c....t.....ac.........cacc.aac..a.c..tcg...a....a..a..a.g....c.
H. lar          ................t.c.....c.........a....t....tc....a.gt.a...........t-.

Cambridge       aaacccaa-tccacatcaaaaccccctccc-catgcttacaagcaagta-c-agcaatcaaccctcaacta
94-107 normal   ........-.....................-.................-.-.................
94-107 mutant   ....ga..c..a...a...ct..t..c.a.-g.c.............g.-.t..a.ta.cttaact...gg
Neanderthal     .....t..-..........c.....c...-................c.-.-...........t......g
Nu insertion    ........-................a...-................-.-......t...........g
P. troglodytes  ..t..act-c.............tt.a.t.-.............c.c.-.-.a........t.c.....g
P. paniscus     .....t.ca.......t.........c...-.....a..t.....c.a.-.-.at....g...tc.....g
G. gorilla      ---------------------------------------------------------------------
P. abelii       ....t.c.-a.....ctcg.c.t..aca..-.c-...............-.-cc.cccatg...c.ccacc
H. lar          ccca....-a.gtacat.c..a.ga.c..aa..............c.c.-....catcttgac.....g

Cambridge       tcacacatcaactgca-actcc-aaagccacccctcacccactaggataccaacaaac-ctacccaccctt
94-107 normal   ................-.....-.....................................-..........
94-107 mutant   .a......t..acc..cc....a....aa..tt....a....a..cg......a....a...tat.t....
Neanderthal     ...t.........a..-.....-....a.g...t.a.............t.......-..........
Nu insertion    ................-..c..-.......a...ct.........a...t.....g..-.c.t..c.....
P. troglodytes  ..ga....a..aca..-.t...-..c.a........c.....ccc..........g..-...t.tc.....
P. paniscus     ..ga....a..-ca.c-c.c..-....a...t..c.c.....ccc.............-..ga.agt....
G. gorilla      ------------------------------------....a.g.a...t....t..t-.c.tt.ct.a.a
P. abelii       ca.a.acat.caccg.-t...tcc.cata........a..c.c..c....a.cag.c.-aa..aa...t.a
H. lar          .a.-g....c...-tc-....t-c.c.a---------.at.aaca.ca....gt...g-a..gt--..a.c

Cambridge       aacagtacatagtacataaagccatttaccgtacatagcacat---tac---agtcaaatcccttctcgtc
94-107 normal   ...............................................---...---.........t.........
94-107 mutant   --g.............ct.at.gc.c.tg...g..........---.c.---...t...caag.c......
Neanderthal     g.....................................................---.........t.......c.
Nu insertion    ..t.c.......c.....................................---...---.........tc......
P. troglodytes  g....a.............c.a....ac.................---...---.......c....c....c.
P. paniscus     ............c.....c.att..a.................---...---..........a.c....c.
G. gorilla      ..-.........c........t......t..............---.ct---...t....-.a.c..t.c.
P. abelii       ..--........c.....ctat.c.-a....c...........---cc.-----..t...a....g...a..
H. lar          t.a..gg...ga.g..ctc.tt....c....c.....ca.ac.ccc...cac.c....c..a.aatc.a.a

Cambridge       cccatggatgacccccctcagataggggtcccttga
94-107 normal   ...................................
94-107 mutant   aat.c....-.t.............t.....g...t
Neanderthal     ...............................t
Nu insertion    ....c.....................a.........g
P. troglodytes  ....c.....ct..............aa.......g
P. paniscus     ....c.....c..............aa.......g
G. gorilla      ....c.....c..............a.........
P. abelii       ....c.....c..........t...taa......ac
H. lar          .a.------a..c.att...c..g.aa..tt.cc.c
```

Figure 4.6. Maximum homology alignment of the HV1 region of mtDNA from a number of sources. Cambridge sequence (nt 16023 to nt 16399) is from Anderson *et al.* (1981), 94-107 normal and mutant are from this study. The mutant sequence is a consensus from 6 clones with individual clone sequences shown in Appendix 7. Maximum homology alignment was performed using the EMBL Homology Alignment Program in DNASIS. The neanderthal sequence is from Krings *et al.* (1997). The Nu (nuclear) insertion is from Zischler *et al.* (1995) and aligns with the Cambridge sequence between nt 16089 and nt 16399. The *Pan troglodytes*, *P. paniscus*, *Gorilla gorilla*, *Pongo pygmaeus abelii* and *H. lar* sequences are from EMBL and Genbank (accession nos. D38113, G643684, X93347, X97707 and X99256). k = g or t. Dots(.) indicate synonymous nucleotides to the Cambridge sequence. Dashes (-) are present for alignment purposes only, but may indicate possible sites of insertions and deletions. Note that the large deletion in the middle of the *G. gorilla* sequence actually contains the following non-alignable sequence (based on Foran *et al.*, 1988):5′ acaacccccccttcccccccccctcctcca 3′

Analysis of further sample 94-107 amplicons

Presently, all of the mutant DNA is contained in non-contiguous amplicons. To test whether all of the consensus mutations are on the same strand of DNA one can, for example, amplify long stretches of DNA between the facing primers of the HV1 and ND4/ND5 amplicons, clone the PCR product, and then analyse by SSCP and sequencing to see if the ends of the fragments in individual clones are mutant or normal. Unfortunately, as with Southern analysis for testing the pseudogene hypothesis, 'long' PCR requires relatively high concentrations of highly purified template DNA and this was unobtainable because of a lack of semen sample. An analysis of the changes found in the COII, ND4/ND5 and HV1 amplicons (Table 4.4) can also help to show if all pseudogenes are contiguous. GC content, transition/transversion ratios, and homology with other species were all taken into consideration.

Table 4.4. Comparison of GC content, transition/transversion ratios and species homology for sequenced regions of 94-107. For the similarity analysis, the first number refers to the mutant, the second to the Cambridge sequence. *G. gorilla* HV1 sequence only includes the alignable region. For the % changes explained in the HV1 column the brackets enclose the absolute number of transitions and transversions explained. Note that the percent of changes explained by human sequence uses MITOMAP (Wallace *et al.*, 1995) and so includes a large amount of human RFLP and sequence data from which to draw polymorphisms. *P. troglodytes*, *P. paniscus* and *G. gorilla* comparisons are made with just one sequence from each of these species (ref. Fig. 4.6) even though these species are assumed to have as much between individual variation as humans (eg. Goldberg and Ruvolo, 1997).

	COII	ND4/ND5	HV1
% 94-107 mutant GC content	45	43	40
% Cambridge GC content	46	44	46
% AG/GA transitions	29	46	13
% TC/CT transitions	71	46	34
% transversions	0	8	53
% similarity to Cambridge	98	95	72
% similarity to *P. troglodytes*	92/90	90/91	67/83
% similarity to *P. paniscus*	93/91	93/92	67/81
% similarity to *G. gorilla*	88/87	90/90	60/83
% changes explained by Human	0	21	33(31/4)
% changes explained by *P. troglodytes*	86	58	19(17/3)
% changes explained by *P. paniscus*	86	58	18(17/2)
% changes explained by *G. gorilla*	57	52	21(12/1)

94-107 COII amplicon amplification

There is a consensus of seven differences between the mutant and normal strands of the sequenced region (nt 7610 to nt 7975) of the COII amplicon (see Chapter 3 for details). A number of mutant clones had additional changes that were seen only once. One clone, although not sequenced in its entirety, did not have a G → A at nt 7757 but did have flanking mutations. Given that it is highly likely that the changes seen are a result of amplification of a nuclear pseudogene, it is also likely that this missing base is the result of allelic forms of this pseudogene. Neither Davis *et al.* (1997), Hirano *et al.* (1997) nor Wallace *et al.* (1997) found a G → A at nt 7757 within the COII pseudogene. Mosaics could additionally be explained by *Taq* polymerase strand switching (see previously) but in this instance this would require two switching events on the same molecule. The SSCP screening showed just one clear extra band associated with the mutation, indicating that, if intermediates are present, that they are in very low proportions. SSCP analysis of 65 clones shows that the clone ratio between normal and mutant is 1 : 0.5. In the present study no apparent heteroplasmy was evident in other regions of the COII amplicon although changes were found throughout COII in other studies (Hirano *et al.*, 1997; Wallace *et al.*, 1997). No heteroplasmy was detected from the COI region in this study. Possible reasons for this are discussed below.

From Table 4.4 it is seen that the 'mutant' COII sequence does not have a markedly altered GC ratio (although six of the seven changes result in the loss of a C or G) probably because of the relatively small number of mutations. The i/v ratio is skewed towards transitions (with no transversions) which is normal for mtDNA (Brown *et al.*, 1982). Whilst the mutant sequence has more homology with humans than any other species, primate sequences have marginally more homology with the mutant sequence than with humans, with most of the changes being explained by single primate sequences.

94-107 ND4/ND5 amplicon amplification

The ND4/ND5 heteroplasmy was initially detected as *Alu*1 RFLPs. Forty eight consensus changes were observed out of a sequenced region of 890 bp. Of additional interest, are instances where not all of the changes were present in all clones. Of the 10 clones

analysed, four had a mixture of mutant and normal DNA. Three of these were normal at one end and mutant at the other, and the fourth (178-1-531-10) was only missing the T → C transition at nt 11770 (Appendix 6). Bradley and Hillis (1997) experimentally demonstrated that *Taq* polymerase produced recombinant sequences three out of seven times in a 1128 bp amplicon. Theoretically, the rate of strand switching is dependent on amplicon length, polymerase processivity and fidelity, and PCR extension time but these factors have not been experimentally confirmed.

*Alu*1 digests of ND4/ND5 amplicon from 94-107 show that a number of *Alu*1 cut sites are themselves variably heteroplasmic. Sites at nt 12214 and nt 12284 were cut to produce 531 bp and 601 bp mutant fragments, but occasionally failed to be cut, producing 890 bp mutant fragments. Variable mutations at nt 12218 and nt 12285 could account for these results. The lack of any normal *Alu*1 143 bp fragment suggests, as mentioned above, that the normal sequence is polymorphic at the *Alu*1 site at nt 11683. The mutant strands were cut here suggesting that they were not polymorphic for this site.

As for COII, the mutant ND4/ND5 sequence does not have a markedly altered GC ratio (Table 4.4). The i/v ratio is skewed towards transitions (92% transitions) which is normal for mtDNA (Brown *et al.*, 1982). The mutant sequence has more homology to humans than any other species, but with more than 50% of the changes being explained by single primate sequences. As with the COII sequence, it is practically impossible to predict the effect that these mutations would have on protein structure and function given the large number of changes. The obvious exception to this is the new premature termination signal caused by a change at nt 12441. This would create a severely truncated ND5 polypeptide (34mer vs. 603mer), obviously altering protein structure and function. None of the tRNA changes affect the anticodon or amino acid acceptors, but beyond this it is difficult to predict the effects of the changes on tRNA. Given that a severely truncated ND5 protein would alter mitochondrial complex 1 activity and the 'mutant' ND4/ND5 sequence has homology with primates, it is predicted that this sequence is also a nuclear pseudogene. Whether this pseudogene is related physically and temporally to the COII pseudogene is unknown.

94-107 HV1 amplicon amplification

At least two different types of HV1 DNA is present in semen sample 94-107. The normal type has just one change (C → T at nt 16354) with respect to the Cambridge sequence, and probably defines the donors maternal lineage as North Eastern European in origin. A consensus sequence of the 'mutant' type has 106 transition/transversion mutations, 3 single base deletions, 1 double base deletion and 6 single base insertions compared to both the Cambridge sequence and the other 94-107 HV1 type. Within the 'mutant' type there is within-sequence variation at two positions. Nucleotide 16090 alternates between being a normal T and a mutant G; whilst at nt 16328, one clone (180-9-25) is missing the consensus C → G transversion. Other than these instances, all the clones analysed have all of the changes suggesting that other more divergent intermediate forms are rare or non-existent. Whilst the HV1 region is non-coding, Sbisà *et al.* (1997) recognised a relatively conserved putative hairpin domain (ETAS1) that may be a recognition site for the arrest of heavy strand synthesis. A number of changes in the 'mutant' HV1 sequence are in conserved nucleotides in this putative hairpin. The large sequence divergence and possibility of ETAS1 hair pin disruption suggest that this 'mutant' HV1 sequence is also a nuclear pseudogene. An analysis of sequence characteristics shows that those for HV1 are very different from COII and ND4/ND5 (Table 4.4). The large number of changes and transversion bias has decreased the GC content by 6 percent. Primate sequences are more related to the normal human sequence than to the mutant, with less homology, the longer ago the primate diverged. Individual primate sequences can explain very few of the changes, and when broken down into transitions and transversions it can be seen that transitions can be explained much more frequently than transversions.

Time of insertion for sample 94-107 pseudogenes

The insertion of mtDNA into the nucleus is a frequent phenomenon that occurs in all plants and animals (for a review see Zhang and Hewitt, 1996). The ancient transfer of functional genes to the nucleus has allowed mitochondrial energy production to be under dual control (Wallace, 1986). The different codon usages between nuclear and mtDNA mean that more recent transfers result in pseudogene formation. Mitochondrial DNA

fragments can insert into functional nuclear genes with pathological consequences (Shay and Werbin, 1992). Nuclear insertions may involve non-continuous mtDNA regions combining to form a continuous nuclear pseudogene (Kamimura *et al.*, 1989), or just a single mtDNA region inserting into the nuclear DNA (Zischler *et al.*, 1995).

Both nuclear genes and mtDNA accumulate mutations over time. Coding regions in both the nuclear and mtDNA have similar mutational constraints with respect to the need to retain structure and function of the encoded protein product. Pseudogenes, however, having no known function, have no such constraints. The rate of mutation fixation in pseudogenes is therefore assumed to be higher than that of coding nuclear DNA. Both coding and noncoding regions of mtDNA, however, generally have a higher rate of mutation fixation than nuclear pseudogenes (see Chaper 1). Estimates of this rate difference vary from ten-fold to forty-fold (Zischler *et al.*, 1995). Arctander (1995) has traced one pseudogene insertion and its paralogous mitochondrial gene (Cyt b) through multiple species of South American birds. He found an overall thirty nine-fold difference in the rate of mutation fixation between the nuclear and mitochondrial genes. Collura and Stewart (1995) found a *P. pygmaeus* Cyt b pseudogene, also present in *Hylobates lar* and *H. sapiens*, that evolves at 0.15% per base per year, about 10 times slower than mitochondrial Cyt b. Because of the generally lower rate of change of nuclear pseudogenes compared to the paralogous mitochondrial genes many people believe that these pseudogenes can be used as molecular fossils (Arctander, 1995; Zischler *et al.*, 1995; Blanchard and Schmidt, 1996; Zhang and Hewitt, 1996; Wallace *et al.*, 1997). That is, pseudogenes closely resemble the ancestral sequence at the time of pseudogene insertion. Assuming that all genetic transfer is vertical, pseudogenes can therefore be used to ascribe phylogenetic relationships. However, not all pseudogenes evolve more slowly than their paralogous mtDNA genes (Lopez *et al.*, 1997). The faster rate of domestic cat 12s rRNA pseudogene evolution than mitochondrial 12s rRNA evolution is probably due to the constraints on the mitochondrial gene. Because of this, the use of pseudogenes as molecular fossils must be treated with some caution (Zhang and Hewitt, 1997).

Calculating the time of pseudogene insertion is a complex matter. For recently inserted pseudogenes, it can be assumed that the pseudogene divergence between

individuals is minimal and hence the rate of change since insertion is effectively zero. This was shown experimentally by Zischler *et al.* (1995) with the analysis of a human D-loop pseudogene. As not all humans sampled had this D-loop pseudogene it was assumed that the time of insertion was after the earliest radiation of modern humans. The pseudogene sequence from different races was identical, whilst the paralogous D-loop sequences from different races showed the normal level of variation suggesting that the rate of pseudogene mutation since insertion was effectively zero (Zischler *et al.*, 1995). For ancient pseudogenes, an assumption of a zero rate of change since insertion will introduce greater errors with increasing time of divergence. To calculate the divergence (insertion) time of ancient pseudogenes, an assumption (or calculation) of the rate of pseudogene evolution is required. With the rate of pseudogene evolution and the rate of mtDNA evolution, one can estimate the coalescence time of the mtDNA and nuclear sequences and hence the time of insertion (Li *et al.*, 1981; Collura and Stewart, 1995; Sorenson and Fleischer, 1996).

Prior to calculating times of insertion of pseudogenes it is helpful to have a ball park idea of when the insertion took place (so that the appropriate model can be used for the analysis). Relationships between the inserted sequence and paralogous sequences are used to construct a phylogenetic (or cladistic) tree. Given that the data used in tree construction and divergence times is the same these methods are not independent of one another. To do this analysis, the PAUP (3.0s+9) computer program (Swofford, 1991) was used. In tree analysis, relationships between sequences are seen as branches. Minimum relative branch lengths are seen as a conservative estimate of evolutionary distance. Trees were created using a branch and bound algorithm with bootstrap analysis (Swofford, 1991). Bootstrap analysis quantifies the confidence that a particular branch falls in a particular position within a tree. Only the most parsimonious ND4/ND5 (Fig. 4.7) and HV1 (Fig. 4.8) trees are presented as trees for COII and for ND4/ND5 were very similar. Trees are unrooted as suitable sequences were not available as outgroups (in the case of HV1 it is uncertain if any of the primate sequences would predate the mutant sequence and non-primate HV1 sequences have little alignable sequence) and no ancestral states were available. Additionally, within this analysis, character states are unweighted as the estimates of i/v bias vary between nuclear and mtDNA. Notes on the sequences used are included in the figure legends. An HV1 tree including a *G. gorilla*

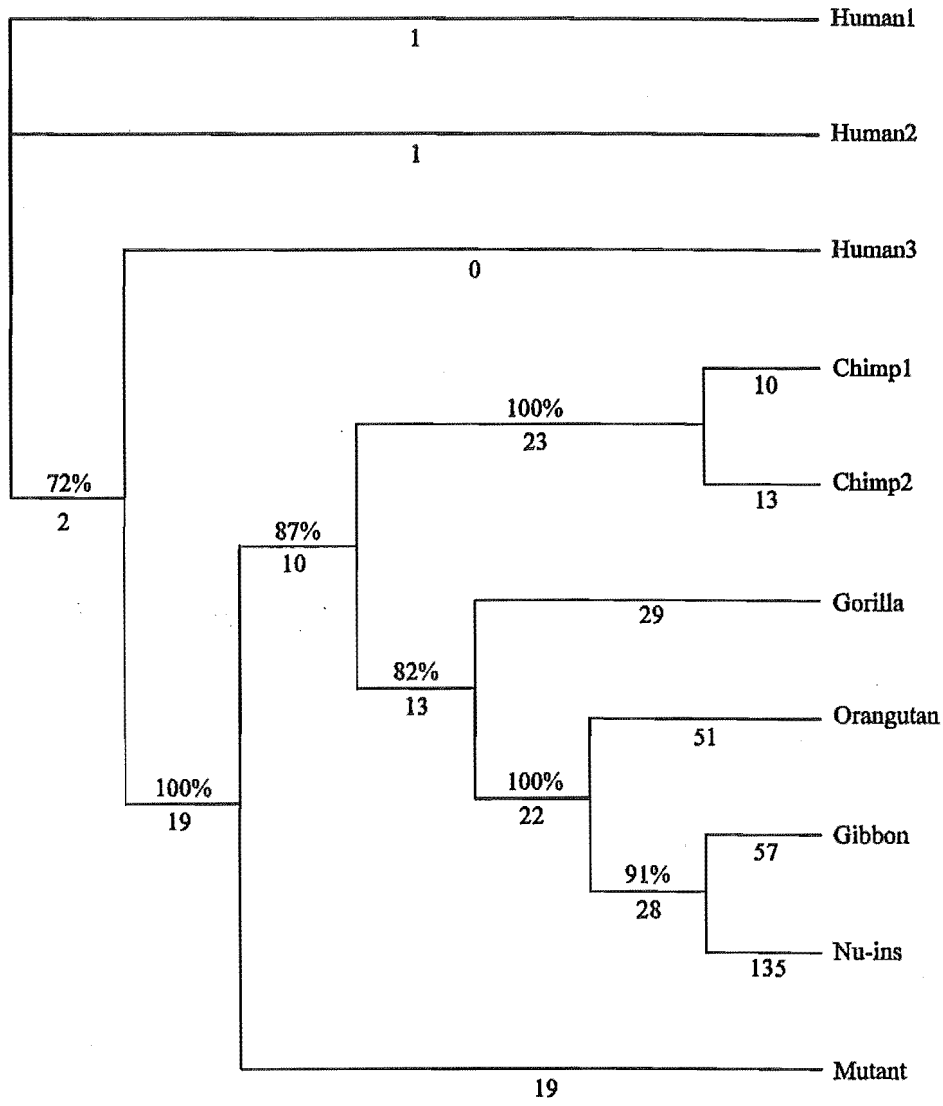


Figure 4.7. Most parsimonious ND4/ND5 sequence phylogenetic tree. Tree was constructed by branch and bound algorithm of PAUP 3.0s+9. Only branches where the bootstrap value was greater than 50% are resolved. Tree length = 580, Consistency index = 0.788, Homoplasy index = 0.212. Branch lengths (below branches) are minimum possible lengths given by PAUP, while bootstrap values (above branches) are calculated from 200 replications. Accession numbers for sequences from EMBL and Genebank are as follows Human1, X62996 (Cambridge sequence); Human2, X93334; Human3, D38112, Chimp2, D38113. Other sequences were derived as follows: Chimp1, Gorilla, Orangutan and Gibbon are from Brown *et al.* (1982); Nuclear insertion (Nu-ins) is clone Lm-P6 from Fukuda *et al.* (1985), whilst mutant is the consensus 'mutant' sequence of sample 94-107 from this study. Sequences were compared over 890 characters (as shown in Appendix 8). Alignment followed Brown *et al.* (1982).

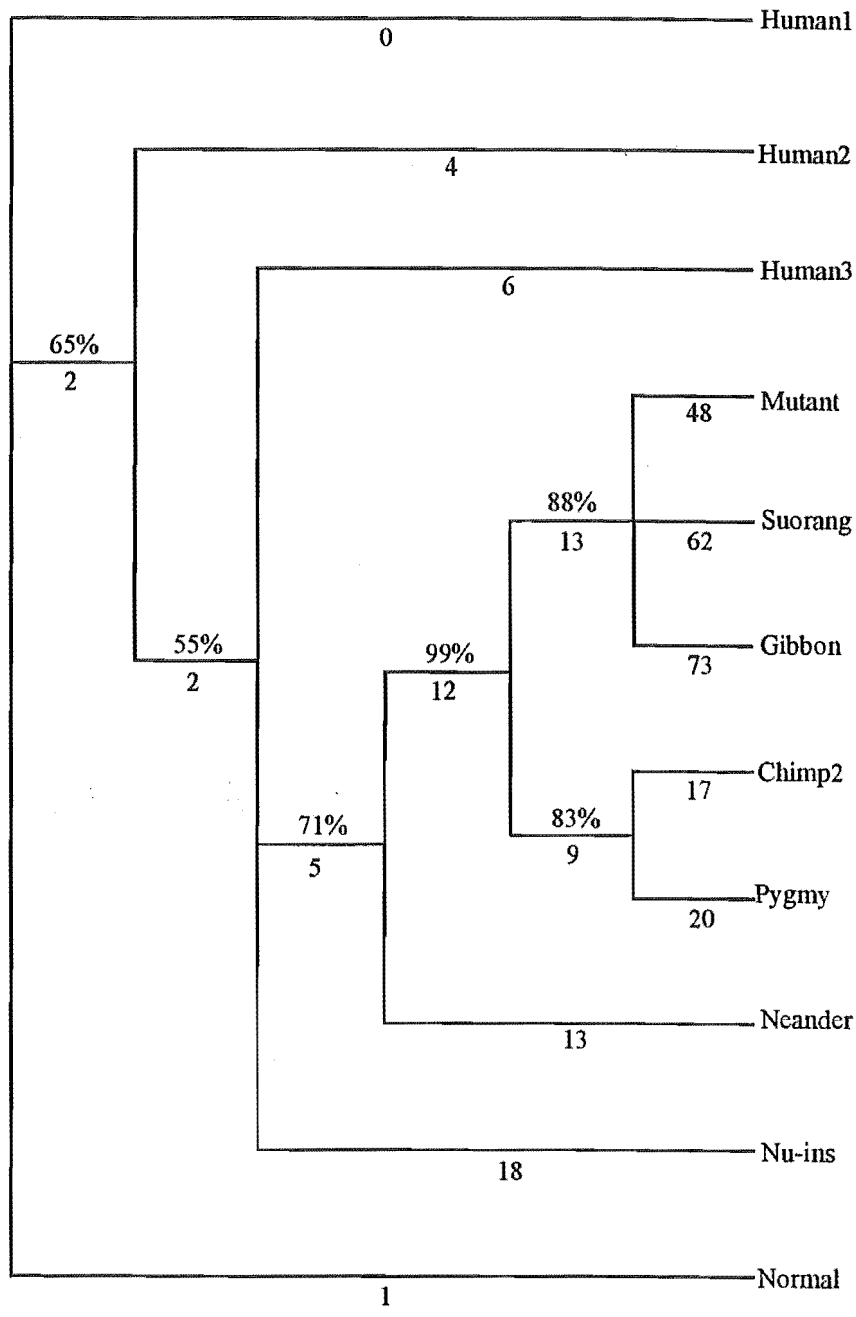


Figure 4.8. Most parsimonious HV1 sequence phylogenetic tree. Tree was constructed by branch and bound algorithm of PAUP 3.0s+9. Only branches where the bootstrap value was greater than 50% are resolved. Tree length = 389, Consistency index = 0.774, Homoplasy index = 0.226. Branch lengths (below branches) are minimum possible lengths given by PAUP, while bootstrap values (above branches) are calculated from 200 replications. Accession numbers for sequences from EMBL and Genebank are as follows Human1, X62996 (Cambridge sequence); Human2, X93334; Human3, D38112; Chimp2, D38113; Pygmy (pygmy chimpanzee *P. paniscus*), D38116, Suorang (Sumatran orangutan, *Pongo pygmaeus abelii*) X97707, Gibbon (*Hylobates lar*) X99256. Other sequences were derived as follows: Normal and mutant are representatives of sample 94-107 from this study, Neander is the consensus neanderthal sequence of Krings *et al.* (1997) and the Nu-ins (nuclear insertion) is derived from Zischler *et al.* (1995). Sequences were compared over 327 characters between nt 16088 and nt 16414 (numbering includes missing values (gaps) included for alignment). Alignment of most sequences was based on the conserved ETAS domains presented at http://www.ba.cnr.it/dloop.html prepared by Sbisà *et al.* (1997). Alignment of the HV1 pseudogene (this study) and *P. pygmaeus abelii* were based on conserved domains at the 3 and 5' ends of HV1 sequence.

sequence used only 175 alignable bases with no branches being resolved with a bootstrap level of 50% or greater. Therefore *G. gorilla* sequences were left off this analysis.

The ND4/ND5 (and COII) trees suggest that the 94-107 ND4/ND5 pseudogene falls between modern human sequences and *P. troglodytes*. The lack of alignable primate sequence and the high sequence diversity of older primates in the HV1 region makes this analysis tenuous. The addition of a large number of further human and *P. troglodytes* sequences to this analysis does not enhance this analysis and so these have been left out of these trees. The small number of differences between (Human1, Human2) in ND4/ND5 and (Human1, Normal) in HV1 makes these branches impossible to resolve at a bootstrap confidence level of 50%. It is anticipated that the saturated sequence divergence of *H. lar*, *P. pygmaeus abelii* and the present study HV1 pseudogene make this trichotomy impossible to resolve above the bootstrap confidence limit of 50%.

From Table 4.4 and the above phylogenetic relationships, it was assumed that the COII and ND4/ND5 pseudogenes were the result of recent insertions but that the HV1 pseudogene was inserted much longer ago. Current theories of hominid evolution (and indeed the above trees) indicate that humans are most closely related to *P. troglodytes*, with *G. gorilla* and other primates being less closely related but more closely related than other mammals. African apes diverged from *P. pygmaeus* 13 million years ago (Pillbeam, 1984). Horai and others (1992) have used this time of divergence to estimate that the time of the last common ancestor of modern humans and *P. troglodytes* was between four and five million years ago. By computing the number of mtDNA nucleotide differences between humans and *P. troglodytes* and correcting for reversion mutations it is possible to estimate a rate of change per million years. By assuming that this molecular clock (evolutionary rate of mutation fixation) is constant across the board one can directly calculate the time of the last common ancestor between two species without there having to be a fossil record. Whilst this is most predictive using coding DNA, due to the constraints on mutation fixation and therefore the relative lack of reversion mutations, this method has been performed on D-loop mtDNA sequences (Vigilant *et al.*, 1991 and refs therein).

Using a three-substitution-site (3ST) formula from Kimura (1981) the probabilities of transitions and transversions per nucleotide between sequences were calculated. Between-sequence transition (TC and AG substitutions) frequencies per site were termed P. Per site, between sequence transversion frequencies were divided into TA and CG substitutions (Q) and TG and CA substitutions (R). These frequencies were then corrected for multiple substitutions at a site using equation (1) (Kimura, 1981; eq. 6) to give the overall corrected substitution rate (K) (rate of fixation at a nucleotide position) as shown in Table 4.5.

$$K = -0.25 \times \ln[(1-2P-2Q)(1-2P-2R)(1-2Q-2R)] \quad (1)$$

If the divergence time (T) is known, then the base substitution rate per unit time (k_{nuc}) is given by equation (2) (derived from Kimura, 1981; eq. 7; see Appendix 9)

$$k_{nuc} = K/(T) \quad (2)$$

Table 4.5. Corrected nucleotide substitution rates between different sequences. Sequences compared are the Cambridge (H), 94-107 'mutant' (M), *P. troglodytes* (C) and *G. gorilla* (G). Chimpanzee and gorilla sequences can be seen in the appropriate appendices. COII n = 365, ND4/ND5 n = 890, HV1 n = 238 where (n) is the number of nucleotides used in the analysis. Note that for HV1-K values, the sequence used is only that over which *G. gorilla* has alignment.

	COII-K	ND4/5-K	HV1-K
M-H	0.01955	0.05681	0.31990
M-C	0.08953	0.08755	0.34935
M-G	0.13988	0.10500	0.37314
H-C	0.10615	0.09590	0.14763
H-G	0.14348	0.11485	0.19787
C-G	0.16140	0.11740	0.26527

Assuming a time of last common ancestor (T) of humans and *P. troglodytes* (H-C) of between four and five million years, the rate of nucleotide substitution per site per million years (k_{nuc}) (calculated for COII) is between 0.0212 and 0.0265. Then assuming that the rate of change since insertion of the COII pseudogene was zero, the theoretical time of insertion of the COII pseudogene (calculated from M-H) is K for M-H divided by the constant rate of substitution per million years calculated above. This gives M-H divergence (insertion) times of between 0.74 - 0.92 million years before present. When the same analysis was done for the ND4/ND5 pseudogene the times of M-H divergence

(pseudogene insertion) were 2.36 - 2.95 million years before present. These estimates are based on a large number of assumptions and have errors of unknown magnitude associated with them. The major assumption is that mtDNA nucleotide substitution rates are constant within and between species over time.

Wallace *et al.* (1997) estimated a time of insertion of 770 thousand years before present for the COII pseudogene which is in the range of 0.74 - 0.92 million years estimated in this study. Wallace and coworkers estimate is based on more substitutions and so is likely to be more accurate. Both estimates place the insertion after the divergence of hominids from the present *P. troglodytes* lineage but before the major human radiation as estimated by Cann *et al.* (1987). Additionally, the estimates of both Wallace *et al.* (1997) and the present study are based on zero rates of change of the pseudogene since insertion. Within sample 94-107, in addition to the changes in the pseudogene found by Wallace *et al.* (1997), there is a G $\rightarrow$ A at nt 7757. This is a mosaic, however, as it is not found in clone 3-1-4. The presence of this change, even within one individual, suggests that the COII pseudogene is not static.

The ND4/ND5 pseudogene has never been reported previously. A divergence time of 2.36 to 2.95 million years places the insertion point after the hominid - *P. troglodytes* divergence but prior to the major human radiations (Horai *et al.*, 1992). It is expected, therefore, that all modern humans would harbour this insertion. Like the COII pseudogene, the ND4/ND5 pseudogene appears to be a mosaic and therefore a zero rate of change may have underestimated the divergence (insertion) time.

The HV1 pseudogene, based on Table 4.4 and the phylogenetic tree analysis (Fig. 4.8) is estimated to have a much older insertion time. It is therefore inappropriate to use the zero change method for calculating the time of insertion. Using the method of Li *et al.* (1981), however, requires the resolution of phylogenetic branches beyond the time of insertion. Within the present study these branches could not be resolved with greater than 50% confidence (Fig. 4.8). Because of this, the previous method was used to give a very conservative estimate of insertion time. Assuming a time of last common ancestor (T) of humans and *P. troglodytes* (H-C) of between four and five million years, the rate of nucleotide substitution per site per million years (k_{nuc}) (calculated for HV1) is

between 0.0295 and 0.0369. Then, assuming that the rate of change since insertion of the HV1 pseudogene is zero, the theoretical time of insertion of the HV1 pseudogene (calculated from M-H) is K for M-H divided by the constant rate of substitution per million years calculated above. This gives M-H divergence (pseudogene insertion) times of between 8.5 - 10.5 million years before present. This is considered to be a relatively inaccurate estimate of insertion time because of the assumption that the nuclear pseudogene mutation rate is zero. Collura and Stewart (1995) using mtDNA Cyt b genes and associated pseudogenes estimated that *H. lar* diverged from the great apes about 16 million years ago. The HV1 pseudogene found within sample 94-107 may or may not predate this *H. lar* divergence.

The HV1 region of the mtDNA D-loop is made up of a number of different domains with differing amounts of sequence conservation between species (Saccone *et al.*, 1991; Sbisà *et al.*, 1997). Sbisà *et al.* (1997) recognise two ETAS domains in all mammals studied. These sequences are relatively conserved and may have roles in the regulation of transcription and replication. In primates the ETAS domains are separated by an insertion sequence (IS). In humans this IS is the most variable stretch of DNA in the entire D-loop (Wallace *et al.*, 1995). Sbisà *et al.* (1997) noted that "in this domain human is more similar to gibbon, which possesses the IS element, than to evolutionally closer species such as gorilla and orangutan." Within the present study an alignment of ETAS domains shows that *H. sapiens*, *P. troglodytes*, *P. paniscus*, *P. pygmaeus abelii* and *H. lar* all share similarity in the IS region whereas *G. gorilla* and *P. pygmaeus* have divergent IS sequences (Fig. 4.9). The HV1 pseudogene IS also shares similarity to the human IS. Given the proposed time of insertion of the HV1 pseudogene in the nuclear DNA, and the fact that pseudogenes may act as molecular fossils, this finding supports the suggestion that the IS seen in *H. lar*, *H. sapiens*, *P. troglodytes*, *P. paniscus* and *P. pygmaeus abelii* is ancestral and that the IS elements of *G. gorilla* and *P. pygmaeus* have diverged from this.

Given the calculated time of insertion of the 94-107 HV1 pseudogene, it may seem surprising that this sequence has not been seen previously. Most studies analysing HV1 variation use PCR primer binding sites directly flanking, and within two base pairs of, the HV1 region; in $tRNA^{Pro}$ on the 5' end of HV1 and in the D-loop central domain

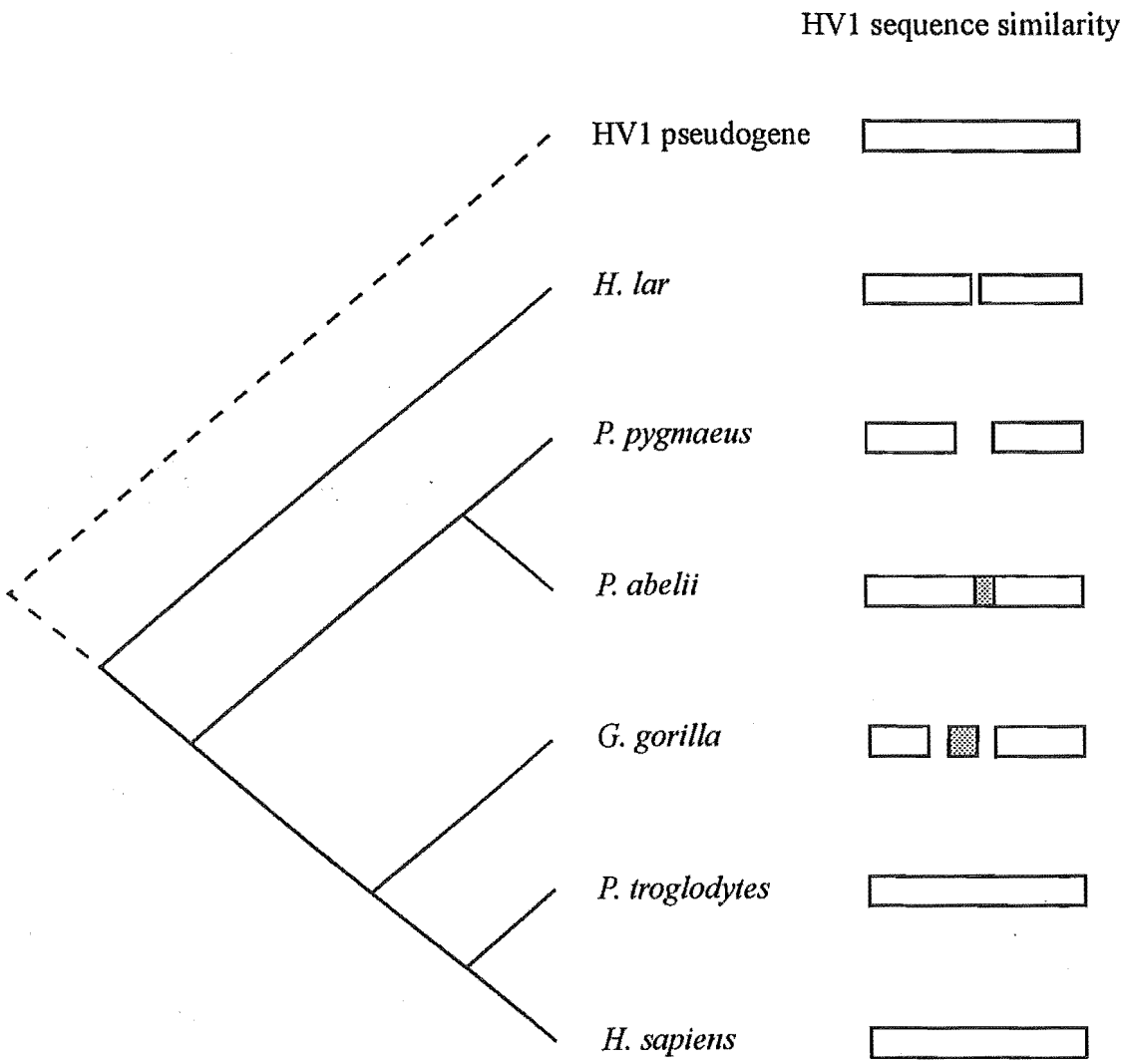


Figure 4.9. Phylogenetic relationship of six primate species and their HV1 alignments with respect to *H. sapiens*. Shaded regions indicate IS sequences with no similarity to *H. sapiens* IS sequence. Sequence gaps are for alignment purposes only but may indicate sites of insertions and deletions. HV1 sequences are all from Figure 4 6. The phylogenetic relationship is as per Collura and Stewart, (1995). The position assignment of the 94-107 HV1 pseudogene branch is tentative.

on the 3' end of HV1 (eg. Vigilant *et al.*, 1991; Krings *et al.*, 1997). The 94-107 HV1 pseudogene, however, includes at least part of $tRNA^{Pro}$. Within $tRNA^{Pro}$ the HV1 pseudogene contains at least a T → C transition at nt 16017 (results not shown). This transition falls within the binding site of the upstream primer of all of the above studies, suggesting these primers would not bind, and the pseudogene would not be amplified. The upstream primer chosen in the present study binds 5' of the conventional upstream primer.

Method analysis

Given the number of pseudogenes characterised in sample 94-107 it may appear surprising that more were not seen in other amplicons. There are a number of possible reasons for this. Firstly, one or both of the PCR primers may fall on a substitution site. This would then mean that the primers will prime only off the normal mtDNA strands and only the mtDNA sequence would be amplified. SSCP will then show only a normal pattern and no RFLPs would be seen unless they also occur in the normal mtDNA sequence. A second possibility is that the nuclear DNA being amplified may be restricted to the pseudogenes found to date. Lastly, it is possible that the SSCP lacks some sensitivity given the lack of changes found in the other *Bst*N1 cleaved COII fragments that contain changes reported by Hirano *et al.* (1997) and Wallace *et al.* (1997).

PCR-SSCP screening of the COII amplicon and analysis of the clone sequences shows that approximately 34% of the DNA is pseudogenic (densitometry of HV1 shows approximately the same levels). Sperm have approximately 50 copies of mtDNA each but are haploid for nuclear DNA. White blood cells within semen are diploid but have 10 fold more copies of mtDNA than sperm. Therefore, the effective ratio of nuclear DNA to mtDNA templates is 1:50 or higher. Primers were designed specifically for mitochondrial DNA, and touchdown PCR was employed to increase priming specificity. Given the template ratio and specificity of priming it was expected that mitochondrial DNA would be preferentially amplified (at least 50 times higher than nuclear DNA) but this was not seen in the observed mutant ratio. Like the study of Davis *et al.* (1997), however, the PCR template in the present study was prepared by boiling cells. When extracting DNA from white blood cells Hirano *et al.* (1997) found that boiling preferentially extracts

nuclear DNA with only partial extraction of mitochondrial DNA (presumably due to the resistance of mitochondrial membranes to boiling). This would severely alter the number of mtDNA templates available for PCR whilst increasing in proportion the number of nuclear templates. This is sufficient to explain the seemingly incongruous normal to 'mutant' ratio.

These pseudogenes were found in only one out of 223 individuals in the present study. Given the calculated dates of pseudogene insertions this is highly unlikely to be the true frequency in any human population. Hiraro *et al.* (1997) found the COII pseudogene in four Alzheimer disease patients and four unselected controls. 94-107 is one of four semen samples in the present study with a sperm count over 300 x 10^6 per ml. MtDNA PCR reactions were not template concentration specific but rather contained an equivalent of 0.5 µl of semen, irrespective of sperm count. The 94-107 COII PCR reaction therefore contained more total DNA than most others in the present study. It is anticipated, therefore that the limiting factor in amplifying these pseudogenes was nuclear DNA concentration within the PCR. To this end, Hirano *et al.* (1997) effectively criticised Davis *et al.* (1997) for using a boiling method for DNA extraction. I still feel justified in using this technique for the reasons given in Chapter 2, and because in almost all instances mtDNA was amplified preferentially to a supposedly common nuclear pseudogene. None of the semen samples with counts over 300 x 10^6 sperm/ml, other than 94-107, had the COII pseudogene amplified.

Chapter 5

GENERAL DISCUSSION

Sensitivity of SSCP analysis to detect mutations in the ATPase and COII amplicons

The sensitivity of SSCP to detect mutations in the present study is unknown. Little base line data is available about the inter-individual variation in the ATPase and COII genes in humans. Hofmann *et al.* (1997) used a PCR-SSCP screening approach but found no polymorphisms in 67 German individuals in the ATPase or COII genes. Tanaka and Ozawa (1994) on the other hand using direct sequencing found 9 COII and 6 ATPase gene polymorphisms in 43 Japanese individuals. Marzuki *et al.* (1991) sequenced mtDNA from 13 unrelated European individuals and found three COII and 10 ATPase gene polymorphisms. The studies of Tanaka and Ozawa (1994) and Marzuki *et al.* (1991) sequenced individuals with known mtDNA diseases whilst that of Hofmann *et al.* (1997) used control individuals.

In this study one heteroplasmic and four different homoplasmic point mutations were found in the ATPase amplicon of 210 semen samples (Table 2.1). Seven different homoplasmic point mutations, one homoplasmic deletion, and one heteroplasmic substitution/insertion were found in the screening of 223 semen samples for variation in a COII amplicon (Table 3.1). Two base substitution heteroplasmies seen in SSCP analysis of the COII amplicon of semen samples AM433B and 94-107 were artefactual.

The ATPase and COII amplicons overlap by 274 bp, the overlap being made up of 101 bp of COII, the COII/$tRNA^{Lys}$ intergenic spacer, $tRNA^{Lys}$ and 77 bp of ATPase 8. COII screening identified three mutations in this ATPase/COII amplicon overlap; a 9 bp deletion, a transition at nt 8251 and a heteroplasmic point mutation and insertion in semen sample TM. These three mutations were seen in eighteen semen samples in the COII amplicon screening. Seventeen of these 18 semen samples were also screened for changes in the ATPase amplicon (Appendix 1). None of the three mutations in the ATPase/COII amplicon overlap seen in COII screening were seen during the ATPase

screening. This lack of SSCP sensitivity on the ATPase amplicon was thought to be caused by the excessive length of the fragments used in the SSCP analysis in the ATPase amplicon. Ways of overcoming this problem (discussed in Chapter 2) were impractical for the present study.

The effect of size of the DNA fragment on SSCP sensitivity is much debated. The optimum fragment size is predicted to be between 100 and 300bp, with fragments of 400 - 500bp having a sensitivity of between 70 and 80% (Beier, 1993). Different DNA fragments of similar sizes can yield mutations to SSCP analysis with very different frequencies suggesting that sequence context is as important as DNA fragment size (Sheffield *et al.*, 1993).

It is considered that the sensitivity of SSCP analysis to detect mutations in the COII amplicon was reasonable although no studies are published that give a good estimation of the expected number of polymorphisms within the COII gene. The work of Davis *et al.* (1997), although flawed in its claims of COII heteroplasmy (see Chapter 4), could provide these data. No polymorphisms from the study of Davis and coworkers have been published.

PCR-SSCP as a means of detecting heteroplasmy

Differentiating between true and false heteroplasmy is very important. The polymerase chain reaction has an inherent error rate resulting in artefactual heteroplasmy. Artefactual heteroplasmy resulting from PCR errors was seen in three instances in this study (results not shown) but was discounted by repeat PCR amplification and screening. PCR primers were chosen so that they would not bind to sequences containing known mutations or polymorphisms.

PCR primer binding is specific but not discriminatory. In semen sample 94-107, the apparent extensive heteroplasmy was created by PCR primers binding to complementary nuclear DNA and mtDNA. The co-amplification of mtDNA and nuclear DNA leading to ambiguous results has been reported previously (Collura and Stewart, 1995; Davis *et al.*, 1997; Hirano *et al.*, 1997; Wallace *et al.*, 1997). To avoid co-amplification of nuclear pseudogenes some studies suggest carrying out specific

enrichment or purification of mtDNA from nuclear DNA (Arctander, 1995; Collura and Stewart, 1995). With small sample volumes this is often not possible. An alternative approach used in this study, is to use small amounts of template DNA in the PCR reaction to eliminate pseudogene amplification by dilution effects. Thus, while the COII pseudogene is likely to be common (Hirano *et al.*, 1997; Wallace *et al.*, 1997), it was only found in one semen sample in the present study. By reducing template concentration one runs the risk of underestimating the real level of heteroplasmy, as rare mtDNA mutants may be subjected to allele drop out. Traditional non-PCR based methods of mtDNA isolation and cloning also negate the problem of nuclear pseudogenes.

An additional cause of false heteroplasmy is sample contamination. To rule out sample contamination, new samples must be analysed, preferably in a different laboratory. While it was not possible to do this with semen sample AM433B, the changes found in heteroplasmy were also found in homoplasmy in semen sample QD788L (Chapter 3) suggesting contamination of the latter with the former.

MtDNA mutations and fertilising potential

From the present study it seems unlikely that mtDNA mutations are a frequent primary cause of poor male fertilising potential although in order to show this definitively, both a larger sample size and more mtDNA genes have to be screened with higher resolution and sensitivity. More and more studies are highlighting the importance of secondary mutations in mtDNA diseases (reviewed in Torroni and Wallace, 1994). Recent evidence suggests that secondary mutations may be seemingly neutral mtDNA polymorphisms (Lertrit *et al.*, 1994). In addition to acting as secondary mutations, polymorphisms (categorised in populations as haplotypes) can define a genetic susceptibility to mtDNA diseases. Hofmann *et al.* (1997) showed that LHON and other mtDNA diseases were more prevalent in certain German haplotypes, even though the disease causing mutations themselves arose spontaneously. Marchington *et al.* (1996) hypothesised that a T $\rightarrow$ C transition at nt 16189 within the D-loop creates a predisposition for the formation or fixation of deleterious mutations in $tRNA^{Leu(UUR)}$. This transition at nt 16189 creates a homopolymeric C_{10} tract followed by length heteroplasmy akin to the T $\rightarrow$ C transition at nt 8827 in semen sample TM in the present study.

Relationship of ATPase mutations to fertilising potential

Only one of the homoplasmic point mutations found in the ATPase amplicon was found in more than one semen sample. This change, causing a *Hae*II RFLP, was not fully characterised (Chapter 2; pattern 4). The proportion of samples that were oligozoospermic and had pattern 4 was not statistically significantly different from the proportion of samples in the sampled population that were oligozoospermic. The same was true for those samples that were asthenozoospermic and had pattern 4. Therefore oligo and asthenozoospermia were not associated with pattern 4.

A T → C transition at nt 8821 was the only heteroplasmic change found in the ATPase amplicon. Though predicted to deleteriously affect ATPase 6 structure and function, it is unknown what effect this change has on the fertilising potential in semen sample QR568B. The repeat sample of QR568B, QF205F, had 100% abnormal head morphology in mature sperm that do not have detectable levels of the mutation. The proportion of mutant DNA required for phenotypic effects in sperm is unknown. Folgerø *et al.* (1993) did not quantify the proportion of the A → G transition heteroplasmy at nt 3243 in the mitochondrial $tRNA^{leu(UUR)}$ found to affect sperm motility in one individual. Nor, in fact, did Folgerø *et al.* (1993) even show that this individual was heteroplasmic although the rest of the individual's maternal family was heteroplasmic for this change (Folgerø *et al.*, 1995). In the present study, the nt 8821 heteroplasmic transition within the ATPase 6 gene was not detected in 36 single mature sperm or mature sperm aggregates so it is not known if it causes a reduction in sperm motility akin to the mutation of Folgerø and co-workers study. The proportion of mutant DNA required for phenotypic effects is probably dependent on the distribution of the mutant DNA in the semen sample and the pathogenicity of the mutation itself. The five other pathogenic mutations seen in the ATPase 6 gene (Chapter 2) all require well above 90% mutant DNA for phenotypic effects in somatic tissues. As such, these ATPase 6 mutations are effectively homoplasmic. The proportion of PCR products from semen samples QR568B and QD205F with the T → C transition at nt 8821 was 16 to 18%. The mutant DNA from semen samples QR568B and QD205F was detected, almost in homoplasmy, from two aggregates of immature sperm. Whether the T → C transition at nt 8821 had any bearing on the lack of maturation of these aggregates is not known. This transition has not been seen in heteroplasmy or homoplasmy previously and the amino acid residue that nt 8821 helps to encode is invariant within vertebrates (Fig. 2.9). Given that the heteroplasmic T → C at nt

8821 was not found in mature sperm it may have a secondary affect on the fertilising potential of QR568B and QD205F's donor by reducing the number of mature sperm within the ejaculate.

Relationship of COII mutations to fertilising potential

No true coding region heteroplasmies were found in the COII amplicon. Semen sample TM had heteroplasmic length variation within the COII/tRNA^{Lys} spacer. Homoplasmic changes have been seen in this spacer before as neutral length variants (Chapter 3). There is no reason to suspect the TM length variant is anything other than a neutral variant.

The G → A transition at nt 8251 (Chapter 3; pattern 3) was found in the COII amplicon from 11 semen samples. Two out of the 10 samples with pattern 3, for which data were available, were oligozoospermic and two were asthenozoospermic. These proportions were not significantly different to the proportions of oligozoospermic and asthenozoospermic semen samples in the sampled population. Therefore the presence of oligo and/or asthenozoospermia was not associated with the COII amplicon pattern 3. The other homoplasmic change in the COII amplicon, found in more than one sample was a 9 bp deletion in the COII/tRNA^{Lys} intergenic spacer (Chapter 3; pattern 2). Up to 94% of Polynesian individuals in New Zealand carry this deletion (Sykes *et al.*, 1995). In the present study it is likely that the donors of all 7 semen samples with this deletion have a Polynesian maternal ancestry. Twenty eight percent of the sampled population were oligozoospermic, while 33% of the six semen samples with pattern 2 were oligozoospermic (excluding missing data). These proportions are not statistically significantly different. The five semen samples, for which sperm motility was available, that had the 9 bp deletion were all asthenozoospermic. This proportion of asthenozoospermia is statistically significantly different to the 46% of asthenozoospermic samples in the sampled population. This significance was nullified when other donor characteristics were considered. Two of the 7 semen samples with this deletion were from the same individual who had Kallmann's syndrome, likely to cause severe seminal defects. It is unlikely that the 9 bp deletion has a pathological linkage with asthenozoospermia, but that the relationship is based on cultural differences. Men in Polynesian ethnic groups may be less likely to seek help with mild male infertility

problems. Additionally, men in Polynesian ethnic groups may be less likely to give consent for the residue of their semen ejaculate to be used for research purposes unless their fertility problem was severe. As shown below, Polynesian men are under-represented in the sampled population.

Was the sampled population representative of the Christchurch population ?

Eighty six percent of residents in the Christchurch urban area identify with a European ethnicity, 9% with a Polynesian ethnicity and four percent with an Asian ethnicity (Statistics New Zealand, 1997). Ethnicity is self perceived and so individuals can belong to more than one ethnic group. The uncharacterised *Hae*II RFLP creating ATPase amplicon pattern 4, is a marker for European haplotype K (Torroni *et al.*, 1991) and is found in 7.4% of European individuals. This RFLP was seen in 13 semen samples (6.2%) in the present study which is not significantly different to the European population proportion at the 95% confidence limit. The G $\rightarrow$ A transition at nt 8251 within the COII amplicon causes a *Hae*III RFLP. While this marker is not only found in Europeans it has been used to define European haplotype I. Between 7.4% (Torroni *et al.*, 1994) and 7.5% (Shoffner *et al.*, 1993) of Europeans have haplotype I. Within this study, 4.9% of the samples contained this change (11/223). This is not statistically significantly different to the population percentage of between 7.4 and 7.5%. Seven semen samples from six individuals (3.1%) had a 9 bp deletion in the COII/tRNA^{Lys} intergenic spacer. The proportion of semen samples with the 9 bp deletion in the sampled population was statistically significantly different to the proportion of the Christchurch population that identifies with a Polynesian ethnicity ($Chi^2 = 9.28$, $df = 1$, $P > 0.05$). This implies that Polynesian men are under-represented in the sampled population which may support the idea that cultural differences are responsible for the linkage between the presence of the 9 bp deletion and asthenozoospermia.

Nine homoplasmic single base substitutions in the ATPase and COII amplicons were seen from individual semen samples. None of these are predicted to have pathogenic consequences, hence all are assumed to be population polymorphisms. All single base substitutions found only once were present at a population frequency of less than 0.5%. None of these changes have been defined previously as haplotype markers with a defined population proportion.

Two true heteroplasmies were observed in the present study. Although neither the substitution heteroplasmy in semen sample QR568B nor the length heteroplasmy in semen sample TM were shown to be confined to mtDNA, this was inferred. Single cell and aggregate PCR-SSCP analysis of the T → C transition at nt 8821 within semen sample QR568B showed that both almost-homoplasmic normal and mutant SSCP patterns could be obtained from cell aggregates. If the mutation were a nuclear insertion it is expected that all single cells and aggregates analysed would appear heteroplasmic. The homopolymeric tract length heteroplasmy in the COII/$tRNA^{Lys}$ intergenic spacer of semen sample TM is an example of a multiplasmy. For this to be the result of pseudogene amplification, multiple insertions of different length homopolymeric mutations would have to be present in semen sample TM.

All mtDNA mutations have to go through a heteroplasmic phase. An analysis of the true mutation rate of mtDNA must incorporate both the rate of change fixation into the population and the rate of unfixed heteroplasmy (Parsons *et al.*, 1997). Direct sequencing of mtDNA may not resolve unfixed heteroplasmy because heteroplasmy may be indistinguishable from sequencing artefacts. Mutation detection systems such as SSCP analysis and the Heteroduplex method have the potential to detect all heteroplasmy with a proportion of mutant DNA above 5%, and are limited only by their mutation detection sensitivity. That an appreciable level of true heteroplasmy was not detected in this study could either be because the true level of heteroplasmy was low or that the SSCP analysis lacked sensitivity. Alternatively, the choice of tissue may have affected the level of heteroplasmy detected (Kiechle *et al.*, 1996).

Future directions

Improvements in experimental approach

If this study were to be repeated there are a number of possible improvements to the experimental approach taken. SSCP is a laborious, and potentially insensitive approach to mutation detection. Capillary electrophoresis based methods have the potential to increase speed and accuracy (Mitchelson *et al.*, 1997) but detection of DNA molecules relies on laser induced flourescence making this a method with high setup costs. The recent advent of oligomer-chip technology (Hoheisel, 1997) and its

application to mtDNA mutation analysis (Chee *et al.*, 1996) makes this an attractive future possibility. Oligomer-chip technology is presently limited by cost and an inability to detect heteroplasmy.

To definatively show that a mutation causes disease one has to directly demonstrate a biochemical defect and an additional lack of other confounding mutations. Neither of these two things were done in this study. To demonstrate a biochemical defect one would have to measure the respiratory activity of semen samples or of stable lymphoblastoid cultures derived from donated blood. To demonstrate that no other mutations were present, one would have to sequence firstly, the whole mtDNA genome, and secondly, the genes of all nuclear derived polypeptides associated with OXPHOS energy production.

Fertility analysis

One of the problems in determining if mtDNA mutations are associated with fertilising potential is the lack of quantitative data. In order to get more quantitative data a very large multicentre population study could be undertaken. An alternative to this may be available.

In 1992, Carlsen *et al.* gleaned from the literature 61 reports of seminal counts and volume from 1938 to 1990. Statistical analysis of data from these reports showed that over the past 50 years, human sperm counts in first world countries have dropped by an average of 42%, along with a decrease in seminal volume. Opponents of the decrease point out that literature selection bias and variability of collection methods make such a trend statistically insignificant (Olsen *et al.*, 1995). Opponents have also pointed out that literature pertaining to the past 20 years show mean sperm counts increasing. Laboratories in Finland (Vierula *et al.*, 1996) and the USA (Macleod and Wang, 1979, cited in Carlsen *et al.*, 1992) both have records showing no change in seminal counts over the 50 year period, although the mean seminal counts are different. Similarly, the seminal counts from farm animals have shown no change over the same period (Setchell, 1997). What these studies have highlighted, however, is that there are geographical differences in sperm count irrespective of whether there has been a temporal decline. The fact that Finnish men have higher sperm counts than English men, for example, has been

attributed to different lifestyles, with suggestions being put forward that men in urban centres have lower sperm counts than men in rural areas (eg. Bujan *et al.*, 1996 cf. Auger *et al.*, 1995) and that those on organic diets having higher sperm counts than those who are not (Abell *et al.*, 1994). Whatever the reason, mtDNA polymorphisms (and hence haplotypes), given that they are population specific, may act as markers for this variability in sperm count and/or may act as susceptibility factors for the decrease in sperm count over the past 50 years. In this way the issue of variability of mtDNA and its interaction with fertilising potential could be addressed.

Pseudogene analysis

The apparent heteroplasmy in the COII amplicon of semen sample 94-107 was created by the co-amplification of a nuclear pseudogene. Further analysis using PCR primers specific for mtDNA identified two more pseudogenes co-amplified with mtDNA. The COII and ND4/ND5 pseudogenes have sequences intermediate between human and *P. troglodytes* suggesting that the times of pseudogene insertion were since the time of the last common ancestor of these two species. The HV1 pseudogene, however, is very divergent from all human and primate sequences. The mtDNA D-loop has a very high substitution rate (Parsons *et al.*, 1997) and therefore any potentially useful phylogenetic information is rapidly obscured by multiple substitutions at single nucleotide positions. No HV1 regions with the same IS structure have been found that can be accurately be predated to the insertion of the HV1 pseudogene from semen sample 94-107. The accurate assignment of insertion times using the model of Li *et al.* (1981), however, requires the resolution of phylogenetic branches beyond the time of insertion. Because of this it has not been possible to date the time of HV1 pseudogene insertion with any degree of accuracy. The HV1 pseudogene boundaries are presently defined by the site of primer binding. It is anticipated that the inserted sequence was larger than just HV1, however. The 5' flanking genes of HV1 are $tRNA^{Pro}$, $tRNA^{Thr}$ and Cyt b. These genes are highly conserved between mammalian species. If paralogous pseudogenes of these genes were part of the HV1 pseudogene, they could be sequenced as more informative markers of insertion time.

ACKNOWLEDGMENTS

I thank my supervisor, Associate Professor Frank Sin for his support, encouragement, guidance and for generally giving me a free rein to follow my ideas.

I thank my associate supervisors, Dr Iris Sin and Dr Peter Benny for arranging access to semen samples and for ironing out ethical hurdles.

Thank you to all of the hundreds of anonymous semen donors who made this and many more studies possible.

Thank you to Mary Whyte, Jan Deans and Tina Newsome from the Andrology Laboratory, Department of Pathology, Canterbury Health for invaluable semen analysis.

A big thank you to those people who have communicated unpublished ideas and results including Dr Anne Jaquier, Dr Lyndall Hatch, Dr Bob Lightowlers, Dr Tom Parsons, Dr Joanna Poulton and Dr Maryellen Ruvolo.

Thanks to Dr Mike Steel for helping with phylogenetic analysis, and especially for sorting out the equations in Chapter 4.

Many thanks to the Technicians and Secretaries of the Zoology Department who have fielded questions and solved problems with ease.

Thanks to all the people I have had the pleasure of working with in the lab including Mel, Sylvia, Seumas, Ujjal, Kirstie, Logan, John, Simon, Sue, Fiona, Nat and Caro.

Thanks to Dr Jenny Khoo (aka lab mum) for not growling too much when I did stupid things and for teaching me most of what I know.

Thanks to the New Zealand Health Research Council for a Seeding Grant to initiate this investigation.

Thanks to the University of Canterbury for providing me with a Doctoral Scholarship so that I could undertake this project, and thanks to the Canterbury Branch of the Royal Society of New Zealand for providing me with a travel award so that I could communicate it.

Thanks to Johnny for the late night discussions, computer games and proofing poorly written drafts.

Thank you to my parents for encouraging me to pursue my dreams and introducing me to science from a very young age.

Finally, a huge thank you to Louise for her constant encouragement, love and support for the duration of this thesis. Thanks for dating me at the start, marrying me in the middle and not divorcing me at the end of this degree !

REFERENCES

Abell, A., Ernst, E. and Bonde, J.P. (1994) High sperm density among members of organic farmers' association. *The Lancet* **343**:1498.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson,A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., and Young, I.G. (1981) Sequence and organisation of the human mitochondrial genome. *Nature* **290**:457-465.

Anderson, S., Bankier, A.T., Barrell, B.G., deBruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1982) Comparison of the human and bovine mitochondrial genomes. In: *Mitochondrial Genes* (eds: Slonimski, P., Borst, P. and Attardi, G.) Cold Spring Harbour, New York.

Anderson, W.A. and Perotti, M.E. (1975) An ultrachemical study of the respiratory potency, integrity and fate of the sea urchin sperm mitochondria during early embryogenesis. *Journal of Cell Biology* **66**:367-376.

Arctander, P. (1995) Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proceedings of the Royal Society of London, Series B* **262**:13-19.

Arnason, U., Gullberg, A. and Janke, A. (1997) Phylogenetic analyses of mitochondrial DNA suggest a sister group relationship between Xenarthra (Edentata) and Ferungulates. *Molecular Biology and Evolution* **14**(7):762-768.

Arnason, U., Gullberg, A. and Widegren, B. (1993) Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. *Molecular Biology and Evolution* **10**(5):960-970.

Auger, J., Kunstmann, J.M., Czyglik, F. and Jouannet, P. (1995) Decline in sperm quality among fertile men in Paris during the past 20 years. *New England Journal of Medicine* **332**:281-285.

Avise, J.C. (1991) Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual Review of Genetics* **25**:45-69.

Ballinger, S.W., Schurr, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K-H. and Wallace, D.C. (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* **130**:139-152.

Bandy, B. and Davison, A.J. (1990) Mitochondrial mutations may increase oxidative stress: implications for carcinogenesis and aging. *Free Radical Biology and Medicine* **8**:523-539.

Barrientos, A., Casademont, J., Solans, A., Moral, P., Cardellach, F., Urbano-Márquez, A., Estivill, X. and Nunes, V. (1995) The 9-bp deletion in region V of mitochondrial DNA: evidence of mutation recurrence. *Human Genetics* **96**:225-228.

Bassam, B.J., Caetano-Anolles, G. and Gresshoff, P.M. (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analytical Biochemistry* **196**:80-83.

Beier, D.R. (1993) Single-strand conformation polymorphism (SSCP) analysis as a tool for genetic mapping. *Mammalian Genome* **4**:627-631.

Bendall, K.E., Macaulay, V.A., Baker, J.R. and Sykes, B.C. (1996) Heteroplasmic point mutations in the human mtDNA control region. *American Journal of Human Genetics* **59**:1276-1287.

Bendall, K.E. and Sykes, B.C. (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *American Journal of Human Genetics* **57**:248-256.

Bibb, M.J., van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) Sequence and gene organisation of mouse mitochondrial DNA. *Cell* **26**:167-180.

Blanchard, J.L. and Schmidt, G.W. (1996) Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Molecular Biology and Evolution* **13**(3):537-548.

Boulet, L., Karpati, G. and Shoubridge, E.A. (1992) Distribution and threshold expression of the tRNALys mutation in skeletal muscle of patients with myoclonic epilepsy and ragged-red fibres (MERRF). *American Journal of Human Genetics* **51**:1187-1200.

Bradley, R.D. and Hillis, D.M. (1997) Recombinant DNA sequences generated by PCR amplification. *Molecular Biology and Evolution* **14**(5):592-593.

Brown, M.D. and Wallace, D.C. (1994) Molecular basis of mitochondrial DNA disease. *Journal of Bioenergetics and Biomembranes* **26**(3):273-289.

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *Journal of Molecular Evolution* **18**:225-239.

Bujan, L., Mansat, A., Pontonnier, F. and Mieusset, R. (1996) Time series analysis of sperm concentrations in fertile men in Toulouse, France between 1977 and 1992. *British Medical Journal* **312**(7029):471-472.

Calafell, F., Underhill, P., Tolun, A., Angelicheva, D. and Kalaydjeva, L. (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Annals of Human Genetics* **60**:35-49

Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) Mitochondrial DNA and human evolution. *Nature* **325**:31-36.

Cann, R.L. and Wilson, A.C. (1983) Length mutations in human mitochondrial DNA. *Genetics* **104**:699-711.

Cantore, P., Roberti, M., Rainaldi, G., Gadaleta, M.N. and Saccone, C. (1989) The complete nucleotide sequence, gene organisation, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. *The Journal of Biological Chemistry* **264**(19):10965-10975.

Capaldi, R.A. (1990) Structure and function of Cytochrome *c* Oxidase. *Annual Review of Biochemistry* **59**:569-596.

Cavalli-Sforza, L.I. (1998) The DNA revolution in population genetics. *Trends in Genetics* **14**(2):60-65.

Carlsen, E., Giwercman, A., Keiding, N. and Skakkebæk, N.E. (1992) Evidence for decreasing quality of semen during past 50 years. *British Medical Journal* **305**:609-613.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P.A. (1996) Accessing genetic information with high-density DNA arrays. *Science* **274**:610-614.

Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S. and Wallace, D.C. (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all continent-specific haplogroups. *American Journal of Human Genetics* **57**(1):133-149.

Chou, P.Y. and Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology* **47**:45-148.

Collura, R.V. and Sewart, C-B. (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* **378**:485-489.

Comas, D., Pääbo, S. and Bertranpetit, J. (1995) Heteroplasmy in the control region of human mitochondrial DNA. *PCR Methods and Applications* **5**:89-90.

Cortopassi, G.A., Shibata, D., Soong, N-W. and Arnheim, N. (1992) A pattern of accumulation of a somatic deletion of mitochondrial DNA in aging human tissues. *Proceedings of the National Academy of Sciences, USA* **89**:7370-7374.

Cotton, D.G.H. (1989) Detection of single base changes in nucleic acids. *Biochemical Journal* **263**:1-10.

Cox, G.B., Fimmel, A.L., Gibson, F. and Hatch, L. (1986) The mechanism of ATP synthase: a reassessment of the functions of the b and a subunits. *Biochimica et Biophysica Acta* **849**:62-69.

Cui, X., Li, H., Goradia, T.M., Lange, K., Kazazian, H.H., Galas, G. and Arnheim, N. (1989) Single sperm typing: determination of genetic distance between the $^{G}\gamma$-globin and parathyroid hormone loci by using the polymerase chain reaction and allele specific oligomers. *Proceedings of the National Academy of Sciences, USA* **86**:9389-9393.

Cummins, J. (1997) Mitochondrial DNA: implications for the genetics of human male infertility. In: *Genetics of Male Fertility*. (eds: Barratt, C., de Jonge, C., Mortimer, D. and Parinaud, J.) EDK Editions, Paris, France (in press).

Cummins, J.M., Jequier, A.M. and Kan, R. (1994) Molecular biology of human male infertility: links with aging, mitochondrial genetics, and oxidative stress? *Molecular Reproduction and Development* **37**:345-362.

Cummins, J.M., Meloni, B.P. and Jaquier, A.M. (1993) Human sperm mitochondrial DNA: PCR amplification. In: *Australian Society for Reproductive Biology 25th Annual Conference*, Dunedin, NZ. (Abstract)

Davis, R.E., Miller, S., Herrnstadt, C., Ghosh, S.S., Fahy, E., Shinobu, L.A., Galasko, D, Thal, L.J., Beal, M.F., Howell, N. and Parker, W.D.Jr. (1997) Mutations in mitochondrial Cytochrome *c* Oxidase genes segregate with late-onset alzheimer disease. *Proceedings of the National Academy of Sciences, USA* **94**:4526-4531.

de Lamirande, E. and Gagnon, C. (1992) Reactive oxygen species and human spermatozoa II. Depletion of adenosine triphosphate plays an important role in the inhibition of sperm motility. *Journal of Andrology* **13**(5):379-386.

de Vries, D.D., van Engelen, B.G.M., Gabreels, F.J.M., Ruitenbeek, W. and van Oost, B.A. (1993) A second missence mutation in the mitochondrial ATPase 6 gene in Leigh's syndrome. *Annals of Neurology* **34**:410-412.

de Meirleir, L., Seneca, S. and Lissens, W. (1994) Bilateral striatal necrosis with a novel point mutation in the mitochondrial ATPase 6 gene. *Annals of Neurology* **36**:522 (Abstract).

Desjardins, P. and Morais, R. (1990) Sequence and gene organisation of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *Journal of Molecular Biology* **212**:599-634.

Dianzini, I., Camaschella, C., Ponzone, A. and Cotton, R.G.H. (1993) Dilemmas and progress in mutation detection. *Trends in Genetics* **9**(12):403-406.

Fischer, S.G. and Lerman, L.S. (1983) DNA fragments differing by single base-pair substitutions in denaturing gradient gels: correspondence with melting theory. *Proceedings of the National Academy of Sciences, USA* **80**:1579-1583.

Folgerø, T., Bertheussen, K., Lindal, S., Torbergsen, T. and Øian, P. (1993) Mitochondrial disease and reduced sperm motility. *Human Reproduction* **8**(11):1863-1868.

Folgerø, T., Torbergsen, T. and Øian, P. (1995) The 3243 MELAS mutation in a pedigree with MERFF. *European Neurology* **35**:168-171.

Foran, D.R., Hixson, J.R. and Brown, W.M. (1988) Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop synthesis. *Nucleic Acids Research* **16**(13):5841-5861.

Ford, W.C.L. and Harrison, A. (1981) The role of oxidative phosphorylation in the generation of ATP in human spermatozoa. *Journal of Reproduction and Fertility* **63**:271-278.

Frank, S.A. and Hurst, L.D. (1996) Mitochondria and male disease. *Nature* **383**:224.

Fukuda, M., Wakasugi, S., Tsuzuki, T., Nomiyama, H. and Shimada, K. (1985) Mitochondrial DNA-like sequences in the human nuclear genome: characterisation and implications in the evolution of mitochondrial DNA. *Journal of Molecular Biology* **186**:257-266.

Gadaleta, G., Pepe, G., De Candia, G., Quagliariello, C., Sbisa, E. and Saccone, C. (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *Journal of Molecular Evolution 28*:497-516.

Gibbs, R.A., Nguyen, P.N. and Caskey, C.T. (1989) Detection of single DNA base differences by competitive oligonucleotide priming. *Nucleic Acids Research* **17**(7):2437-2448.

Gill, P., Jeffreys, A.J. and Werrett, D.J. (1985) Forensic application of DNA 'fingerprints'. *Nature* **318**:577-579.

Glover, T.D., Barratt, C.L.R., Tyler, J.P.P. and Hennessey, J.F. (1990) *Human male infertility and semen analysis*. Academic Press, London.

Goldberg, T.L. and Ruvolo, M. (1997) The geographic apportionment of mitochondrial genetic diversity in East African Chimpanzees, *Pan troglodytes schweinfuthii*. *Molecular Biology and Evolution* **14**(9):976-984.

Gomer, R., Datta, S. and Firel, R. (1985) Sequencing homopolymer regions *FOCUS* 7(1):6-7.

Gyllensten, U., Wharton, D., Josefsson, A. and Wilson, A.C. (1991) Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**:255-257.

Harlid, A., Janke, A. and Arnason, U. (1997) The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. *Molecular Biology and Evolution*. **14**(7):754-761.

Hartzog, P.E. and Cain, B.D. (1993) Mutagenic analysis of the a subunit of the F_0F_1 ATP synthase in *Escherichia coli*: Gln-252 through Tyr-263. *Journal of Bacteriology* **175**:1337-1343.

Hatefi, Y. (1993) ATP synthesis in mitochondria. *European Journal of Biochemistry*. **218**:759-767.

Hauswirth, W.H. and Laipis, P.J. (1982) Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. *Proceedings of the National Academy of Sciences, USA* **79**:4686-4690.

Hayashi, K. (1992) PCR-SSCP: a method for detection of mutations. *GATA* **9**:73-79.

Hekman, C., Tomich, J.M. and Hatefi, Y. (1991) Mitochondrial ATP synthase complex: membrane topography and stoichiometry of the F_0 subunits. *The Journal of Biological Chemistry* **266**(21):13564-13571.

Hillier, L. and Green, P. (1991) OSP: a computer program for choosing PCR and DNA sequencing primers. *PCR Methods and Applications* **1**:124-128.

Hirano, M., Shtilbans, A., Mayeux, R., Davidson, M.M. DiMauro, S., Knowles, J.A. and Schon, E.A. (1997). Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normal humans due to PCR amplification of nucleus-embedded mtDNA pseudogenes. *Proceedings of the National Academy of Sciences, USA* **94**:14894-14899.

Hofmann, S., Jaksch, M., Bezold, R., Mertens, S., Aholt, S., Paprotta, A. and Gerbitz, K-S. (1997) Population genetics and disease susceptibility: characterisation of central European haplogroups by mtDNA gene mutations, correlation with D-loop varients and association with disease. *Human Molecular Genetics* **6**(11):1835-1846.

Hoheisel, J.D. (1997) Oligomer-chip technology. *Trends in Biotechnology* **15**:465-469.

Holt, I.J., Harding, A.E., Petty, R.K.H. and Morgan-Hughes, J.S. (1990) A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *American Journal of Human Genetics* **46**:428-433.

Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S. and Takahata, N. (1992) Man's place in hominidea revealed by mitochondrial DNA geneology. *Journal of Molecular Evolution* **35**:32-43.

Houshmand, M., Larsson, N.G., Holme, E., Oldfors, A., Tulinius, M.H. and Andersen, O. (1994) Automatic sequencing of mitochondrial tRNA genes in patients with mitochondrial encephalomyopathy. *Biochimica et Biophysica Acta* **1226**(1):49-55.

Howell, N., Halvorson, S., Kubacka, I., McCullough, D.A., Bindoff, L.A. and Turnbull, D.M. (1992) Mitochondrial gene segregation in mammals: is the bottle neck always - narrow? *Human Genetics* **90**:117-120.

Howell, N., Kubacka, I., Halvorson, S., Howell, B., McCullough, D.A. and Mackey, D.A. (1995) Phylogenetic analysis of the mitochondrial genomes from Leber hereditary optic neuropathy pedigrees. *Genetics* **140**(1):285-302.

Howell, N., Kubacka, I. Halvorson, S. and Mackey, D.A. (1993) Leber's hereditary optic neuropathy: the aetiological role of a mutation in the mitochondrial Cytochrome *b* gene. *Genetics* **133**:133-136.

Howell, N., Kubacka, I. and Mackey, D.A. (1996) How rapidly does the human mitochondrial genome evolve ? *American Journal of Human Genetics* **59**:501-509.

Howitt, S.M., Cleeter, M., Hatch, L. and Cox, G.B. (1993) Functional stability of the a-subunit of the F_0F_1-ATPase from *Escherichia coli* is affected by mutations in three proline residues. *Biochimica et Biophysica Acta* **1144**:17-21.

Huang, C.C., Chen, R.S., Chen, C.M., Wang, H.S., Lee, C.C., Pang, C.Y., Hsu, H.S., Lee, H.C. and Wei, Y.H. (1994) MELAS syndrome with mitochondrial tRNA leu (UUR) gene mutation in a Chinese family. *Journal of Neurology, Neurosurgery, and Psychiatry* **57**:586-587.

Jazin, E.E., Cavelier, L., Eriksson, I., Oreland, L. and Gyllensten, U. (1996) Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. *Proceedings of the National Academy of Sciences, USA* **93**:12382-12387.

Jenuth, J.P., Peterson, A.C., Fu, K. and Shoubridge, E.A. (1996) Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nature Genetics* **14**:146-151.

Johns, D.R., Rutledge, S.L., Stine, O.C. and Hurko, O. (1989) Directly repeated sequences associated with pathogenic mitochondrial DNA deletions. *Proceedings of the National Academy of Sciences, USA* **86**:8059-8062.

Juvonen, V., Yilkki, J., Aula, P., Nikoslelainen, E. and Savontaus, M-L. (1993) Re-evaluation of the linkage of an optic atrophy susceptibility gene to X-chromosomal markers in Finnish families with Leber hereditary optic neuretinopathy (LHON). *American Journal of Human Genetics* **53**:289-293.

Kamimura, N., Ishii, S., Liandong, M. and Shay, J.W. (1989) Three separate mitochondrial DNA sequences are contiguous in human genomic DNA. *Journal of Molecular Biology* **210**:703-707.

Kaneda, H., Hayashi, J-I., Takahama, S., Taya, C., Lindahl, K.F. and Yonekawa, H. (1995) Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *Proceedings of the National Academy of Sciences, USA* **92**:4542-4546.

Kao, S-H., Chao, H-T. and Wei, Y-H. (1995) Mitochondrial deoxyribonucleic acid 4977-bp deletion is associated with diminished fertility and motility of human sperm. *Biology of Reproduction* **52**:729-736.

Kieche, F.L., Kaul, K.K. and Farkas, D.H. (1996) Mitochondrial disorders: methods and specimen selection for diagnostic molecular pathology. *Archives of Pathology and Laboratory Medicine* **120**:597-603.

Kimura, M. (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences, USA* **78**(1):454-458.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X. and Wilson, A.C. (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences, USA* **86**:6196-6200.

Koehler, C.M., Lindberg, G.L., Brown, D.R., Beitz, D.C., Freeman, A.E., Mayfield, J.E. and Myers, A.M. (1991) Replacement of bovine mitochondrial DNA by a sequence variant within one generation. *Genetics* **129**:247-255.

Kolman, C.J., Sambuughin, N. and Bermingham, E. (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142**:1321-1334.

Kondo, R., Horai, S., Satta, Y. and Takahata, N. (1993) Evolution of hominoid mitochondrial DNA with special reference to the silent substitution rate over the genome. *Journal of Molecular Evolution* **36**:517-531.

Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M. and Pääbo, S. (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* **90**:19-30.

Lamminen, T., Majander, A., Juvonen, V., Wikström, M., Aula, P., Nikoskelainen, E., and Savontaus, M-J. (1995) A mitochondrial mutation at nu 9101 in the ATP synthase 6 gene associated with deficient oxidative phosphorylation in a family with Leber hereditary optic neuroretinopathy. *American Journal of Human Genetics* **56**:1238-1240.

Larsson, N.G., Tulinius, M.H., Holme, E., Oldfors, A., Andersen, O., Wahlstrom, J. and Aasly, J. (1992) Segregation and manifestations of the mtDNA tRNA(Lys) A → G (8344) mutation of myoclonus epilepsy and ragged-red fibres (MERRF) syndrome. *American Journal of Human Genetics* **51**:1201-1212.

Lee, H-C., Pang, C-Y., Hsu, H-S. and Wei, Y-H. (1994) Differential accumulations of 4977 bp deletion in mitochondrial DNA of various tissues in human ageing. *Biochimica et Biophysica Acta* **1226**:37-43.

Lench, N., Stanier, P. and Williamson, R. (1988) Simple non-invasive method to obtain DNA for gene analysis. *The Lancet* **18**:1356-1358.

Lertrit, P., Kapsa, R.M.I., Jean-Francois, M.J.B., Thyagarajan, D., Noer, A.S., Marzuki, S. and Byrne, E. (1994) Mitochondrial DNA polymorphism in disease: a possible contributor to respiratory dysfunction. *Human Molecular Genetics* **3**(11):1973-1981.

Lestienne, P. (1992) Mitochondrial DNA mutations in human diseases: a review. *Biochimie* **74**:123-130.

Lessa, E.P. and Applebaum, G. (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology* **2**:119-129.

Li, H; Cui, X; Arnheim, N. (1990) Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction. *Proceedings of the National Academy of Sciences, USA* **87**:4580 4584.

Li, H; Gyllensten, U.B; Cui, X; Saiki, R.K; Erlich, H.A; Arnheim, N. (1988) Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **335**:414-417.

Li, W-S., Gojobori, T. and Nei, M. (1981) Pseudogenes as a paradigm of neutral evolution. *Nature* **292**:237-239.

Liapis, P.J., Van De Walle, M.J. and Hauswirth, W.W. (1988) Unequal partitioning of bovine mitochondrial genotypes among siblings. *Proceedings of the National Academy of Sciences, USA* **85**:8107-8110.

Lien, S., Kaminski, S., Aleström, P. and Rogne, S. (1993) A simple and powerful method for linkage analysis by amplification of DNA from single sperm cells. *Genomics* **16**:41-44.

Lightowlers, R.N., Chinnery, P.F., Turnbull, D.M. and Howell, N. (1997) Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends in Genetics* **13**(11):450-455.

Lopez, J.V., Culver, M., Stephens, J.C., Johnson, W.E. and O'Brien, S.J. (1997) Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. *Molecular Biology and Evolution* **14**(3):277-286.

Magoulas, A. and Zouros, E. (1993) Restriction-site heteroplasmy in anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. *Molecular Biology and Evolution* **10**(2):319-325.

Marchington, D.R., Hartshorne, G.M., Barlow, D. and Poulton, J. (1997) Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: support for a genetic bottle neck. *American Journal of Human Genetics* **60**:408-416.

Marchington, D.R., Poulton, J., Sellar, A. and Holt, I.J. (1996) Do sequence varients in the major non-coding region of the mitochondrial genome influence mitochondrial mutations associated with disease? *Human Molecular Genetics* **5**:473-479.

Marzuki, S., Noer, A.S., Lertrit, P., Thyagarajan, D., Kapsa, R., Utthanaphol, P. and Byrne, E. (1991) Normal varients of human mitochondrial DNA and translation products: the building of a reference data base. *Human Genetics* **88**:139-145.

Matthews, P.M., Brown, R.M., Morten, K., Marchington, D., Poulton, J. and Brown, G. (1995) Intracellular heteroplasmy for disease associated point mutations in mtDNA: implications for disease expression and evidence for mitotic segregation of heteroplasmic units of mtDNA. *Human Genetics* **96**:261-268.

Meirelles, F.V. and Smith, L.C. (1997) Mitochondrial genotype segregation in a mouse heteroplasmic lineage produced by embryonic karyoplast transplantation. *Genetics* **145**:445-451.

Miller, K.W.P., Dawson, J.L. and Hagelberg, E. (1996) A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. *International Journal of Legal Medicine* **109**:113 (also available at http://www.sscf.ucsb.edu/~kmiller.html).

Mitchelson, K.R., Cheng, J. and Kricka, L.J. (1997) The use of capillary electrophoresis for point-mutation screening. *Trends in Biotechnology* **15**:448-458.

Monnat, R.J. and Loeb, L.A. (1985) Nucleotide sequence preservation of human mitochondrial DNA. *Proceedings of the National Academy of Sciences, USA* **82**:2895-2899.

Montoya, J., Ojala, D. and Attardi, G. (1981) Distinctive features of the 5'-terminal sequences of the human mitochondrial mRNAs. *Nature* **290**:465-470.

Myers, R.M., Larin, Z. and Maniatis, T. (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**:1242-1246.

Nakamura, M., Nakano, S., Goto, Y., Ozawa, M., Nagahama, Y., Fukuyama, H., Akiguchi, I., Kaji, R. and Kimura, J. (1995) A novel point mutation in the mitochondrial $tRNA^{Ser(UCN)}$ gene detected in a family with MERFF/MELAS overlap syndrome. *Biochemical and Biophysical Research Communications* **214**(1):86-93.

Noer, A.S., Sudoya, H., Lertrit, P., Thyagarajan, D., Utthanaphol, P., Kapsa, R., Byrne, E. and Marzuki, S. (1991) A tRNA(lys) mutation in the mtDNA is the causal genetic lesion underlying myoclonic epilepsy and ragged-red fibre (MERFF) syndrome. *American Journal of Human Genetics* **49**:715-722.

Ojala, D., Montoya, J. and Attardi, G. (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**:470-474.

Olsen, G.W., Bodner, K.M., Ramlow, J.M., Ross, C.E. and Lipshultz, L.I. (1995) Have sperm counts been reduced 50 percent in 50 years? A statistical model revisited. *Fertility and Sterility* **63**(4)887-893.

Oostra, R-J., Kemp, S., Bolhuis, P.A. and Bleeker-Wagemakers, E.M. (1996) No evidence for 'skewed' inactivation of the X-chromosome as cause of Leber's hereditary optic neuropathy in female carriers. *Human Genetics* **97**:500-505.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences, USA* **86**:2766-2770.

Ozawa, T., Tanaka. M., Ino, H., Ohno, K., Sano, K., Wada, Y., Yoneda, M., Tanno, Y., Miyatake, T., Tanaka, T., Itoyama, S., Ikebe, S., Hattori, N. and Mizuno, Y. (1991) Distinct clustering of point mutations in mitochondrial DNA among patients with mitochondrial encephalomyopathies and with Parkinson's disease. *Biochemical and Biophysical Research Communications* **176**(2):934-946.

Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P. and Holland, M. M. (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* **15**:363-368.

Penn, A.M.W., Lee, J.W.K., Thuillier, P., Wagner, M., Maclure, K.M., Menard, M.R., Hall, L.D. and Kennaway, N.G. (1992) MELAS syndrome with mitochondrial $tRNA^{Leu(UUR)}$ mutation: correlation of clinical state, nerve conduction, and muscle ^{31}P magnetic resonance spectroscopy during treatment with nicotinamide and riboflavin. *Neurology* **42**:2147-2152.

Petri, B., von Haeseler, A. and Pääbo, S. (1996) Extreme sequence heteroplasmy in bat mitochondrial DNA. *Biological Chemistry* **377**:661-667.

Piercy, R., Sullivan, K.M., Benson, N. and Gill, P. (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *International Journal of Legal Medicine* **106**:85-90.

Pillbeam, D. (1984) The descent of hominoids and hominids. *Scientific American* **250**:84

Poon, M-C., Anand, S., Fraser, B.M., Hoar, D.I. and Sinclair, G.D. (1993) Haemophilia B carrier determination based on family-specific mutation detection by single-strand conformation analysis. *Journal of Laboratory and Clinical Medicine* **122**:55-63.

Poulton, J. (1995) Transmission of mtDNA: cracks in the bottleneck. *American Journal of Human Genetics* **57**:224-226.

Pult, I., Sajantila, A., Georgiev, O., Schaffner, W. and Pääbo, S. (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biological Chemistry Hoppe-Seyler* **375**:837-840.

Ray, P.F., Winston, R.M.L. and Handyside, A.H. (1995) Elimination of allele drop-out in single-cell analysis for diagnosis of cystic fibrosis. *Human Reproduction* **10**(Abstract book 2):64-65. (Abstract).

Reid, F.M., Vernham, G.A. and Jacobs, H.T. (1994) Complete mtDNA sequence of a patient in a maternal pedigree with sensorineural deafness. *Human Molecular Genetics* **3**(8):1435-1436.

Riordan-Eva, P. and Harding, A.E. (1995). Leber's hereditary optic neuropathy: the clinical relevance of different mitochondrial DNA mutations. *Journal of Medical Genetics* **32**:81-87.

Ritte, U., Neufeld, E., Prager, E.M., Gross, M., Hakim, I., Khatib, A. and Bonné-tamir, B. (1993) Mitochondrial DNA affinity of several Jewish communities. *Human Biology* **65**(3):359-385.

Roe, B.A., Ma, D-P., Wilson, R.K. and Wong, J.F.-H. (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *Journal of Biological Chemistry* **260**(17):9759-9774.

Ruvolo, M., Zehr, S., von Dornum, M., Pan, D., Chang, B. and Lin, J. (1993) Mitochondrial COII sequences and modern human origins. *Molecular Biology and Evolution* **10**:1115-1135.

Saccone, C., Pesole, G. and Sbisá, E. (1991) The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *Journal of Molecular Evolution* **33**:83-91.

Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M-L. and Aula, P. (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Research* **5**:42-52.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual* (2 ed). Cold Spring Harbour Laboratory Press, New York.

Sathananthan, A.H., Ratnam, S.S., Ng, S.C., Tarín, J.J., Gianaroli, L. and Trouson, A. (1996) The sperm centriole: its inheritance, replication and perpetuation in early human embryos. *Human Reproduction* **11**(2):345-356.

Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G. and Saccone, C. (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* **205**:125-140.

Seibel, P., Flierl, A., Kottlors, M. and Reichmann, H. (1994) A rapid and sensitive PCR screening method for point mutations associated with mitochondrial encephalomyopathies. *Biochemical and Biophysical Research Communications* **200**(2):938-942.

Senior, A.E. (1990). The proton-translocating ATPase of *Escherichia coli. Annual Review of Biophysics and Biophysical Chemistry* **19**:7-42.

Setchell, B. (1997) Sperm counts in semen of farm animals 1932 - 1995. *International Journal of Andrology* **20**:209-214.

Shadel, G.S. and Clayton, D.A. (1997) Mitochondrial DNA maintenance in vertebrates. *Annual Review of Biochemistry* **66**:409-435.

Shay, J.W. and Werbin, H. (1992) New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. *Mutation Research* **275**:227-235.

Sheffield, V.C., Beck, J.S., Kwitek, A.E., Sandstrom, D.W. and Stone, E.M. (1993) The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* **16**:325-332.

Sherratt, E.J., Thomas, A.W., Gagg, J.W. and Alcolado, J.C. (1996) Nonradioactive characterisation of low-level heteroplasmic mitochondrial DNA mutants by SSCP-PCR enrichment. *BioTechniques* **20**(3):430-432.

Sherry, S.T., Rogers, A.R., Harpending, H., Soodyall, h., Jenkins, T. and Stoneking, M. (1994) Mismatch distributions of mtDNA reveal recent human population expansions. *Human Biology* **66**(5):761-775.

Shoffner, J.M., Brown, M.D., Torroni, A., Lott, M.T., Cabell, M.F., Mirra, S.S., Beal, M.F., Yang, C-C., Gearing, M., Salvo, R., Watts, R.L., Juncos, J.L., Hansen, L.A., Crain, B.J., Fayad, M., Reckord, C.L. and Wallace, D.C. (1993) Mitochondrial DNA varients observed in Alzheimer disease and Parkinson disease patients. *Genomics* **17**:171-184.

Shoffner, J.M., Lott, M.T., Lezza, A.M., Seibel, P., Ballinger, S.W. and Wallace, D.C. (1990) Myoclonic epilepsy and ragged red fibre disease (MERFF) is associated with a mitochondrial DNA tRNA(Lys) mutation *Cell* **61**:931-937.

Shoffner, J.M., Lott, M.T., Voljavec, A.S., Soueidan, S.A., Costigan, D.A. and Wallace, D.C. (1989) Spontaneous Kearns-Sayre/chronic external opthalmoplegia plus syndrome associated with a mitochondrial DNA deletion: A slip-replication model and metabolic therapy. *Proceedings of the National Academy of Sciences, USA* **86**:7952-7956.

Silber, J.R. and Loeb, L.A. (1981) S_1 nuclease does not cleave DNA at single-base mis-matches. *Biochimica et Biophysica Acta* **656**:256-264.

Silvestri, G., Ciafaloni, E., Santorelli, F.M., Shanske, S., Servidei, S., Graf, W.D., Sumi, M. and DiMauro, S. (1993) Clinical features associated with the A → G transition at nucleotide 8344 of mtDNA ("MERFF" mutation). *Neurology* **43**:1200-1206.

Smith, T.A., Whelan, J. and Parry, P.J. (1992) Detection of single-base mutations in a mixed population of cells: a comparison of SSCP and direct sequencing. *GATA* **9**:143-145.

Smith-sørenson, B., Hovig, E., Andersson, B. and Børrenson, A-L. (1992) Screening for mutations in human HPRT cDNA using the polymerase chain reaction (PCR) in combination with constant denaturant gel electrophoresis (CDGE). *Mutation Research* **269**:41-53.

Soodyall, H., Vigilant, L., Hill, A.V., Stoneking, M. and Jenkins, T. (1996) MtDNA control-region sequence variation suggests multiple independent origins of an 'Asian-specific' 9-bp deletion in sub-Saharan Africans. *American Journal of Human Genetics* **58**:595-608.

Sorenson, M.D. and Fleischer, R.C. (1996) Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. *Proceedings of the National Academy of Sciences, USA* **93**:15239-15243.

St John, J.C., Cooke, I.D. and Barratt, C.L.R. (1997) The use of long PCR to detect multiple deletions in the mitochondrial DNA of human testicular tissue from azoospermic and severe oligozoospermic patients. In: *Genetics of Male Fertility*. (eds: C. Barratt, C. De Jonge, D. Mortimer and J, Parinaud) EDK. pp 333-347.

Statistics New Zealand (1997) *Census 96 with supermap 3*. Department of Statistics Te Tari Tatau, Wellington, New Zealand.

Suomalainen, A., Kaukonen, J., Amati, P., Timonen, R., Haltia, M., Weissenbach, J., Zeviani, M., Somer, H. and Peltonen, L. (1995) An autosomal locus predisposing to deletions of mitochondrial DNA. *Nature Genetics* **9**:146-151.

Suzuki, Y., Sekiya, T. and Hayashi, K. (1991) Allele-specific polymerase chain reaction: a method for amplification and sequence determination of a single component among a mixture of sequence variants. *Analytical Biochemistry* **192**:82-84.

Swofford, D.L. (1991) *Phylogenetic analysis using parsimony (PAUP), version 3.0s*. Illinois Natural History Survey, Champaign.

Sykes, B., Leiboff, A., Low-beer, J., Tetzner, S. and Richards, M. (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *American Journal of Human Genetics* **57**:1463-1475.

Tanaka, M. and Ozawa, T. (1994) Strand asymmetry in human mitochondrial DNA mutations. *Genomics* **22**(2):327-335.

Thomas, W.K. and Beckenbach, A.T. (1989) Variation in salmonoid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. *Journal of Molecular Evolution* **29**:233-245.

Thyagarajan, D., Shanske, S., Vazquez-Memije, M., DeVivo, D. and DiMauro, S. (1995) A novel mitochondrial ATPase 6 point mutation in familial bilateral striatal necrosis. *Annals of Neurology* **38**:468-472.

Torroni, A., Lott, M.T., Cabell, M.F., Chen, Y-S., Lavergne, L. and Wallace, D.C. (1994) MtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *American Journal of Human Genetics* **55**:760-776.

Torroni, A. and Wallace, D.C. (1994) Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *Journal of Bioenergetics and Biomembranes.* **26**(3):261-271.

Towner, P. (1993) Recovery of DNA from electrophoresis gels. In: *Essential Molecular Biology, A Practical Approach.* (ed. T.A. Brown), IRL Press, pp 130-131.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yanon, R. and Yosikawa, S. (1995) Structure of metal sites of oxidised bovine heart Cytochrome *c* Oxidase at 2.8 Å. *Science* **269**:1069-1074.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yanon, R. and Yosikawa, S. (1996) The whole structure of the 13-subunit oxidised Cytochrome *c* Oxidase at 2.8 Å. *Science* **272**:1136-1144.

Vierula, M., Niemi, M., Saaranen, M., Saarikoski, S. and Suominen, J. (1996) High and unchanged sperm counts of Finnish men. *International Journal of Andrology* **19**(1):11-17

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. and Wilson, A.C. (1991) African populations and the evolution of human mitochondrial DNA. *Science* **253**:1503-1507.

Vilkki, J., Ott, J., Savontaus, M-L, Aula, P. and. Nikoskelainen, E.K. (1991) Optic atrophy in Leber hereditary optic neuroretinopathy is probably determined by an X-chromosomal gene closely linked to DXS7. *American Journal of Human Genetics* **48**:486-491.

Vilkki, J., Savontaus, M-L. and Nikoskelainen, E.K. (1990) Segregation of mitochondrial genomes in a heteroplasmic lineage with Leber hereditary optic neuroretinopathy. *American Journal of Human Genetics* **47**:95-100.

von Heijne, G. (1991) Proline kinks in transmembrane alpha helices. *Journal of Molecular Biology* **218**:499-503.

Wallace, D.C. (1986) Mitochondrial genes and disease. *Hospital Practice* **21**:77-92.

Wallace, D.C. (1992) Mitochondrial Genetics: a paradigm for aging and degenerative diseases? *Science* **256**:628-632.

Wallace, D.C. (1993). Mitochondrial diseases: genotype versus phenotype. *Trends in Genetics* **9**:128-133.

Wallace, D.C., Lott, M.T., Brown, M.D., Huoponen, K. and Torroni, A. (1995) Report of the committee on human mitochondrial DNA. In: *Human gene mapping 1995: a compendium*. (ed. Cuticchia, A.J.) Johns Hopkins University Press, Baltimore, pp 910-954 (also available at http://www.gen.emory.edu/mitomap.html).

Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M.S., Elsas, L.J. and Nikoskelainen, E.K. (1988) Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* **242**:1427-1430.

Wallace, D.C., Stugard, C., Murdock, D., Schurr, T. and Brown, M.D. (1997) Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. *Proceedings of the National Academy of Sciences, USA* **94**:14900-14905.

Wartell, R.M., Hossieni, S.H. and Moran, C.P. (1990) Detecting base pair substitutions in DNA fragments by temperature gradient gel electrophoresis. *Nucleic Acids Research* **18**:2699-2705.

Wong, T.W. and Clayton, D.A. (1986) DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. *Cell* **45**:817-825.

World Health Organisation (1992) *WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction*. Cambridge University Press, Cambridge.

Wrischnik, L.A., Higuchi, R.G., Stoneking, M., Erlich, H.A., Arnheim, N. and Wilson A.C. (1987) Length mutations in human mtDNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Research* **15**(2):529-542.

Yokogawa, T., Watanabe, Y., Kumazawa, Y., Ueda, T., Hirao, I., Miura, K. and Watanabe, K. (1991) A novel cloverleaf structure found in mammalian mitochondrial tRNA ^{ser}(UCN). *Nucleic Acids Research* **19**(22):6101-6105.

Zhang, D-X. and Hewitt, G.M. (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Evolution and Ecology* **11**(6):247-251.

Zischler, H., Geisert, H., Von Haeseler, A. and Pääbo, S. (1995) A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**:489-492.

Zouros, E., Freeman, K.R., Oberhauser Ball, A. and Pogson, G.H. (1992) Direct evidence for extensive paternal inheritance in the marine mussel *Mytilus*. *Nature* **359**:412-414.

Appendix 1 Seminal data for samples used in this study. Counts are given as million sperm per ml of semen. Motility is a percentage of all sperm. D.O.B is the donors date of birth. ATPase and COII columns denote samples screened in those studies with astricts beside those with a confirmed change. For notes on 'Comments' see the end of this appendix.

	Sample	Count	Motility	D.O.B	ATPase	COII	Comments
1	95-5	24.00	75	1956	y*	y	
2	95-9	25.00	27	1958	y	y	no coagulation
3	95-13	37.00	70	1945	y	y*	
4	95-15	3.10	52	1956	y	y	
5	95-17	5.30	26		y*	y	
6	95-19	15.00	57	1962	y		
7	95-22	6.70	57	1954	y*	y	
8	95-23	12.00	80	1962	y	y	
9	95-24	28.00	28		y	y*	
10	95-26	1.30	36	1960	y	y	
11	95-27	79	66	1954	y	y*	
12	95-29	65.00	36		y	y	AAs
13	95-30	0.10	0		y	y	
14	95-34	7.20	47	1959	y*	y	AAs
15	95-35	23.00	47		y	y	
16	95-37	2.20	32	1961	y	y*	Kallmann's syndrome
17	95-38	84.00	73		y	y	
18	95-40	34.00	62	1958	y	y	no coagulation
19	95-54	26.00	35	1965	y	y	viscous
20	95-56	245.00	69		y	y	viscous
21	95-79	20.00	37		y	y*	
22	94-107	349.00	66	1955	y	y*	
23	94-108	345.00	78	1961	y	y	
24	94-109	210.00	46	1961	y		
25	94-110	446.00	42	1955	y	y	
26	94-124	3.40	75	1961	y	y	from gradient
27	94-149	34.00	40	1961	y*	y	
28	94-149P			1956	y	y	
29	94-150	26.00	19	1951	y	y	
30	94-160	11.00	36	1970	y		
31	94-173	204.00	67		y	y	
32	95-173	125.00	40		y	y	
33	A (SF94-014)	131.00	68	1959	y	y	
34	AM221J	21.00	36	1963	y	y	
35	AM229R	60.00	55	1964	y	y	
36	AM232N	39.60	62	1949	y	y	
37	AM235Q	9.90	32	1958	y		
38	AM241P	112.00	42	1960	y	y	
39	AM264B	0.90	58	1958	y	y	
40	AM295L	16.30	60	1959	y	y	
41	AM302J	17.90	61	1955	y		
42	AM311L	74.00	57	1961	y	y	
43	AM367J	31.30	15	1964	y	y	
44	AM378N	64.00	23	1957	y	y	
45	AM422U	114.00	55	1956	y	y	
46	AM433B	49.00	36	1958	y	y*	from gradient
47	AM435D	2.60	47	1963	y	y	
48	AM436E	70.00	53	1959	y	y*	
49	AM447I	73.00	79	1960	y	y	
50	AM459N	3.10	61	1960	y	y	
51	AM467O	290.00	51	1960	y	y	

52	AM468P	124.00	55	1958	y	y	
53	AM470K	88.30	68	1962	y	y*	
54	AM472M	19.80	57	1949	y	y	
55	AM477R	208	89	1965	y	y	
56	AM498B	22.70	64	1971	y	y	
57	AM539M	54.00	39	1958	y	y	
58	AT440E	173.00	59	1966	y	y	
59	AT449N	35.00	47	1961	y	y	
60	AT470T	33.50	46	1964	y	y	
61	AT635U	148.00	80	1957	y	y	
62	AT648W	21.50	71	1961	y		
63	AT665C	51.00	29	1959	y	y*	abnormal forms
64	AT689M	41.00	81	1957	y	y	
65	AT697N	25.00	73	1963	y	y	
66	AT734U	13.10	29	1966	y	y	
67	AT752B	0.00	0	1959	y	y	
68	B(SF94-015)	48.50	37	1952	y	y	
69	BG	23.00	47	1967	y	y	
70	BT	3.70	31	1961	y	y*	Kallmann's syndrome
71	C (SF94-019)	66.00	62	1963	y	y	
72	CCR	152.00	1		y	y*	viscous
73	CD	0.65	8	1965	y	y	
74	CM	5.20	22		y	y	
75	Collins	0.50	33	1966	y	y	
76	CP	1.60	10		y	y	
77	D (SF94-020)	107.00	53	1955	y	y	
78	DISR	0.00	0	1956	y	y	post vasectomy
79	E(SF94-064)	107.00	53	1961	y	y	
80	EV	7.70	53		y*		no acrosomes
81	EZDW2	19.10	61	1954	y	y	yellowish
82	F(SF94-070)	113.00	57	1964	y		
83	GB1	6.70	63		y*	y	
84	GB2	49.00	7		y	y	
85	GI	61.00	67		y	y	precipitous
86	JH	20.00	87	1959	y	y	
87	JN	15.00	61	1957	y	y*	
88	KT	36.30	3	1967	y	y	
89	MC	5.20	22		y*	y	IgA AAs
90	MR	365.00	59	1961	y	y	
91	NG	19.00	66		y	y	
92	OP992C1	41.00	64	1964		y	
93	OP992C2					y*	
94	PK	88.00	59		y	y	
95	PR1	7.70	53	1959	y	y	
96	PR2	109.00	71		y	y	
97	QD201B	67.00	60	1948	y	y	
98	QD202C	129.00	68	1961	y	y	
99	QD203D	74.00	61	1953	y		
100	QD205F	0.90	17	1968	y*		repeat of QR568B
101	QD216J	151.00	71	1961	y		
102	QD2880	24.00	52		y	y	
103	QD720I	54.00	31		y*	y*	
104	QD741P	8.00	9	1962	y	y	
105	QD742Q	35.00	56	1947	y	y	
106	QD747V	200.00	57	1962	y*	y*	
107	QD762W	196.00	67		y		
108	QD765C	95.00	46	1957	y	y	
109	QD788L	85.00	50	1957	y	y*	
110	QD8070	10.00	7		y*		
111	QD844A	76.00	37	1958	y	y	

112	QD845B	101.00	58	1956	y	y	
113	QD848E	146.00	70	1965	y	y	
114	QD862E	125.00	66	1965		y	
115	QD874	0.05	0	1963	y	y	
116	QH447L	clumpy	0	1927	y		"clumping,yellow"
117	QH448M	59.00	48	1952	y	y	
118	QH474R	93.00	65	1956		y	
119	QH507E	85.00	59	1967	y		
120	QH514E	330.00	48	1961	y	y	
121	QH516G	262.00	61	1966	y	y	
122	QH529M	80.00	51	1962	y	y	
123	QN148Q	13.30	61	1962	y		
124	QN149E	0.00	0	1964	y	y	
125	QN167V	50.00	28	1957	y		
126	QN178C	0.80	34	1966	y	y	
127	QN182W				y	y*	
128	QN187E	6.20	21	1962	y	y	
129	QN188F	207.00	72	1968	y	y	
130	QN191B	220.00	60	1966	y		
131	QN208J	411.00	54	1962		y	
132	QN209K	81.00	67	1955		y	
133	QN213H	7.40	56	1960	y		
134	QN214I	0.15	100	1946	y		
135	QN220H	72.00	43	1965	y		
136	QN222J	44.00	72	1955	y		
137	QN228P	99.00	68	1956	y	y	
138	QN230K	44.00	31	1957	y	y	
139	QN237R	136.00	68	1965	y	y	
140	QN242P	105.00	48	1947	y	y	
141	QN417L	1.90	44	1958	y	y	
142	QR497E	36.40	23	1955	y	y	
143	QR498F	28.60	49	1965	y	y	
144	QR509H	222.00	57	1966	y	y*	
145	QR510B	113.00	56	1963	y	y	
146	QR511C	74.40	55	1962	y*	y	
147	QR522G	0.80	38	1966	y	y	
148	QR568B	0.30	20	1968	y*	y	
149	QR569C	50.60	30	1970	y	y	
150	QU337C	76.30	73	1965	y	y	
151	QU3711	100.00	35		y	y	
152	QU379Q	128.00	70	1953	y		
153	QU380K	3.00	39	1958	y	y*	precipitous
154	QU447K	48.60	85	1967	y	y	
155	QU498F	28.60	49	1965	y	y	
156	QW217C	1.90	44	1958	y	y	
157	RR	6.00	44	1958	y	y	Vas reversal; AAs
158	RS	45.00	20	1965	y	y	
159	SC1	4.60	48	1959	y	y	debris
160	SC2	19.00	63	1959	y	y	
161	SCnew sample	13.50	16		y	y	round cells
162	SF-437	125.00	66	1964		y	
163	SF-459/94-459	204.00	55		y		extra chromosome
164	SF104	65.00	70	1956		y	
165	SF148	27.00	52	1958		y	
166	SF331	38.00	53	1949		y	
167	SF362	24.00	43	1965	y*	y	
168	SF461/94-461	104.00	52			y	
169	SF639	174.00	62			y	
170	SF94-100	41.00	35	1970	y	y	
171	SF94-101	23.30	39	1959	y	y	

172	SF94-102	53.00	40		y	y	
173	SF94-104	21.00	74	1962	y		round cells
174	SF94-114	157.00	45		y	y	
175	SF94-299	7.40	50	1959	y	y*	
176	SF94-303	22.10	6	1959	y	y	
177	SF94-315	47.20	49	1953	y	y	
178	SF94-316	135.00	64		y	y	
179	SF94-344	85.70	27	1961	y	y	
180	SF94-347	82.70	48	1960	y	y	
181	SF94-350	32.00	68	1963	y	y	
182	SF94-356	51.30	72	1961	y	y	
183	SF94-357	201.00	72	1959	y	y	
184	SF94-380	15.80	66	1959	y	y	oily-lubricant?
185	SF94-399	81.00	60	1947	y	y	
186	SF94-408	174.00	78	1964	y	y	
187	SF94-409	118.00	72	1962	y	y*	
188	SF94-69	18.70	38		y	y	
189	SF94-78	51.00	21	1955(?)	y	y	
190	SF94-83	51.60	37		y	y	
191	SF94-86	85.70	35	1959	y*	y	
192	SF94-87	14.60	47		y	y	
193	SF94-90	112.00	74		y	y	
194	SF94-91	21.90	50		y	y	
195	SF94-92	44.00	67		y	y	
196	SF94-97	86.00	72		y	y	
197	SF94-99	4.00	50	1959	y	y	
198	SF95/612					y*	
199	SF95/657	45.00	64	1951		y	
200	SF95-049	105.00	59	1963	y	y	
201	SF95-079#2	107.00	53	1959	y	y	
202	SF95-114	157.00	45		y	y	
203	SF95-120	115.00	62	1958	y	y	
204	SF95-122	26.00	50	1964	y	y*	
205	SF95-138	140.00	74	1964	y	y	
206	SF95-144				y	y	
207	SF95-160	167.00			y	y*	
208	SF95-171	30.00	33	1963	y*	y	
209	SF95-196	16.40	56		y	y	
210	SF95-265	27.00	56	1962	y	y	
211	SF95-555			1955	y	y	
212	SF95-590	80.00	35	1967	y	y	
213	SF95-611	43.00	35	1960	y	y	
214	SF95-628	208.00	37	1961	y	y	
215	SF95-633	48.00	0	1976	y	y*	
216	SF96-150	33.00	55	1965		y	
217	SF96-166	186.00	38	1950		y	
218	SF96-17	154.00	64	1966		y	
219	SF96-173	13.00	36	1960		y	
220	SF96-197	131.00	52	1955		y	
221	SF96-201	87.00	66	1962		y	
222	SF96-202	78.70	64	1958		y	
223	SF96-220	67.00	47	1958		y	
224	SF96-229	75.00	50	1961		y	
225	SF96-238	0.00		1962		y	
226	SF96-255	12.00	65	1961		y	
227	SF96-279	65.00	30	1960		y*	
228	SF96-280	2.30	46	1965		y	cryptorchidism
229	SF96-281	135.00	36	1965		y	
230	SF96-283	8.40	17	1958		y	
231	SF96-289	106.00	60	1962		y	

232	SF96-305	140.00	67	1955		y	
233	SF96-309	1.10	16	1962		y	
234	SF96-310	152.00	56	1955		y	
235	SF96-333	0.50	33	1966		y	
236	SF96-341	126.00	50	1970		y	chemical worker
237	SF96-342	94.00	63	1957		y	
238	SF96-389	0.25	10	1955		y	
239	SF96-393	39.00	57	1964		y	
240	SF96-398	69.00	71	1958		y	
241	SF96-400	62.00	53	1958		y	
242	SF96-434	17.00	26	1964		y	
243	SM	6.40	53			y	
244	STEVENS	81.00	36		y	y	
245	TC	76.00	60	1943	y	y	
246	TI	1.35	22	1951	y*	y	
247	TM	3.40	23	1961	y	y*	cryptorchidism
248	TR				y	y	
249	WT	110.00	26		y		
250	ZC	.05-1	6	1957	y		
251	ZK1	3.30	40	1961	y		
252	ZK2	1.20	43		y		

Notes:
AA's refer to antibodies in the semen that may be present as an immune response to an infection. These can frequently cause clumping in semen. Where known, causes for reduced sperm count and motility are indicated, including vasectomy, cryptorchidism, Kallman's syndrome and the presence of extra chromosomes. Semen from gradients is enriched in motile sperm and so is a biased representative of these samples.

Appendix 2

5% SSCP gel.

3.5 ml Acrylamide mix (49:1)
3.5 ml 5 x TBE
27.5 ml ddH_2O
degas 5 min
0.3 ml 10% Ammonium per sulphate (APS)
11 -14 µl TEMED

Silver stain protocol

Fix for 20 min in 10% Acetic acid and 0.02% Formaldehyde
Wash for 2 min with ddH_2O, repeat twice more
Rinse with 0.02% Sodium thiosulphate for 1 min
Repeat 2 min washes with ddH_2O
Impregnate for 30 min with 0.2% Silver nitrate and 0.03% Formaldehyde
Wash for 20 sec with ddH_2O
Develop with 15 g/L Sodium carbonate, 0.02% Formaldehyde and 0.0005% Sodium thiosulphate
Stop development with 10% Acetic acid

6% Urea PAGE sequencing gel

35 g Urea
10.5 ml Acrylamide (sequencing grade mix; 19:1)
7 ml 10 x TBE
made up to 70 ml with ddH_2O
degas 5 min
0.7 ml 10% APS
11 µl TEMED

prehybridisation solution

6 x SSC
5 x Denharts reagent
0.5% (v/v) SDS
0.5% (w/v) blocking reagent

SOC plates

10 g/l Bactotryptone
10 g/l NaCl
5 g/l Bactoagar
15 g/l Agar
10 ml/l 1M $MgSO_4$
10 ml/l 1M $MgCl_2$
glucose to 20mM

Selection media

Xgal to 20 μg/ml
IPTG to 0.5 mM
Amp to 50 μg/ml

Appendix 3 Single cell data. Only cells or clusters that gave positive PCR results are included. Classes are defined in Table 2.2. Result classes are normal (n) and mutant (m) SSCP patterns.

Date	Cell	Class	Result
1.11.94	1-16	IV	n
1.11.94	17-27	III	n
1.11.94	28-33	II	n
29.11.94	1-3	II	n
29.11.94	5,7,9,11	VI	n
29.11.94	12,15	III	n
29.11.94	13,14	VI	n
29.11.94	16	III	n
29.11.94	17,25	VI	n
29.11.94	18	IV	n
29.11.94	20	V(SRC cluster)	m+n
29.11.94	21-24,26	V(SRC cluster)	n
29.11.94	27	VII(round headed sperm)	n
29.11.94	28	VII(round headed sperm)	m(tiny n)
29.11.94	29	V(no tails)	n
29.11.94	30	V(SRC cluster)	n
29.11.94	31	VI(huge cells)	m+n
29.11.94	34	V(small cells)	m+n
6.12.94	2	V(LRC cluster)	n
6.12.94	4	IV(round head)	n(tiny m)
6.12.94	5	III	n
6.12.94	6,7,9	VII(round headed sperm)	n
6.12.94	11	V(SRC)	n
6.12.94	21	VII(spermatids)	m(tiny n)
14.12.94	2,3,6,9	IV	n
14.12.94	1,10,11	IV(large head)	n
14.12.94	4,5,7	IV(split tail)	n
19.12.94	3,6	IV(large head)	n
6.7.95	1	VII(5-6 immature sperm)	n
6.7.95	2	VII(9-10 immature sperm)	n
6.7.95	3	IV	n
6.7.95	4	VII(mature and immature sperm)	n
6.7.95	5	VII(mature and immature sperm)	m+n
6.7.95	6	VII(mature and immature sperm)	n
6.7.95	7	VII(immature sperm)	n
6.7.95	8	VII(mature sperm, 6 cells)	n
6.7.95	9	VII(immature sperm, 10 cells)	n
6.7.95	10	VII(mature and immature sperm)	n
25.10.95	1,2,6,8	II	n
25.10.95	3	II	m(with high band)
26.10.95	4,6,9	II	n

Appendix 4 Individual COII clone sequences for mutant and normal patterns of semen sample 94-107. Dots(.) indicate synonymous nucleotides to the Cambridge sequence (Anderson *et al.*, 1981). Numbering is based on the Cambridge sequence. Clone 3-1-4 does not have a change at 7757. This clone was only sequenced from 7610 to 7811. Five changes were seen only once and were assumed to be PCR artefact.

```
           7         7         7         7         7         7         7         7
           6         6         6         6         6         6         6         6
           1         2         3         4         5         6         7         8
           0         0         0         0         0         0         0         0
Cambridge  ctacaagacgctacttcccctatcatagaagagcttatcacctttcatgatcacgccctcataatcattttccttat
1-1-43     .....................................t.....................................
1-1-53     .....................................t.....................................
2-1-9      .....................................t.....................................
2-1-20     .....................................t.....................................
2-1-23     .....................................t.....................................
3-1-4      .....................................t.....................................
1-1-42     ........................................................................
2-1-7      ........................................................................

             7         7         7         7         7         7         7         7
             6         7         7         7         7         7         7         7
             9         0         1         2         3         4         5         6
             0         0         0         0         0         0         0         0
Cambridge  ctgcttcctagtcctgtatgccctttcctaacactcacaacaaaactaactaatactaacatctcagacgctcagg
1-1-43     ................c.....................................g.........a......
1-1-53     ................c.........t.............................a.....a......
2-1-9      ................c.....................................a.....a......
2-1-20     ................c.....................................a.....a......
2-1-23     ................c.....................................a.....a......
3-1-4      ................c.....................................a.....a......
1-1-42     ........................................................................
2-1-7      ........................................................................

                7         7         7         7         7         7         7         7
                7         7         7         8         8         8         8         8
                7         8         9         0         1         2         3         4
                0         0         0         0         0         0         0         0
Cambridge  aaatagaaaccgtctgaactatcctgcccgccatcatcctagtcctcatcgccctcccatccctacgcatcctttac
1-1-43     ..........................................t.....................................
1-1-53     ..........................................t.....t.....c.....
2-1-9      ..........................................t.....................................
2-1-20     ..........................................t.....................................
2-1-23     ..........................................t.....................................
3-1-4      ..........................................t.
1-1-42     ...................................c....................................
2-1-7      ........................................................................

                   7         7         7         7         7         7         7
                   8         8         8         8         8         9         9
                   5         6         7         8         9         0         1
                   0         0         0         0         0         0         0
Cambridge  ataacagacgaggtcaacgatccctcccttaccatcaaatcaattggccaccaatggtactgaacctacgagtacac
1-1-43     ...........................t.....................t.....................a.....
1-1-53     ...........................t.....................t.....................a.....
2-1-9      ...........................t.....................t.....................a.....
2-1-20     ...........................t.....................t.....................a.....
2-1-23     ...........................t.....................t.....................a.....
3-1-4
1-1-42     ........................................................................
2-1-7      ........................................................................

             7         7         7         7         7         7
             9         9         9         9         9         9
             2         3         4         5         6         7
             0         0         0         0         0         0
Cambridge  cgactacggcggactaatcttcaactcctacatacttcccccattattcctagaacca
1-1-43     ...........................................................
1-1-53     ...........................................................
2-1-9      ...........................................................
2-1-20     ...........................................................
2-1-23     ...........................................................
3-1-4
1-1-42     ...........................................................
2-1-7      ...........................................................
```

Appendix 5 PCR conditions for primers used in Chaper 4 amplicon analysis (complimentary to Table 4.1).

HMTL321-HMTH455,
94°C/30 sec, 60°C/30 sec, 72°C/1 min 30 sec x5,
94°C/30 sec, 56°C/30 sec, 72°C/1 min 30 sec x5,
94°C/30 sec, 54°C/30 sec, 72°C/1 min 30 sec x25,
2 mM $MgCl_2$.

HMTL604-HMTH735,
94°C/30 sec, 60°C/30 sec, 72°C/1 min 30 sec x5,
94°C/30 sec, 56°C/30 sec, 72°C/1 min 30 sec x5,
94°C/30 sec, 54°C/30 sec, 72°C/1 min 30 sec x25,
2.4 mM $MgCl_2$.

HMTL712-HMTH844,
94°C/30 sec, 60°C/30 sec, 72°C/1 min 20 sec x5,
94°C/30 sec, 56°C/30 sec, 72°C/1 min 20 sec x5,
94°C/30 sec, 54°C/30 sec, 72°C/1 min 20 sec x25,
1.4 mM $MgCl_2$.

HMTL817-HMTH934,
94°C/30 sec, 60°C/30 sec, 72°C/1 min 20 sec x5,
94°C/30 sec, 56°C/30 sec, 72°C/1 min 20 sec x5,
94°C/30 sec, 54°C/30 sec, 72°C/1 min 20 sec x25,
2 mM $MgCl_2$.

HMTL1012-HMTH1246,
94°C/30 sec, 54°C/30 sec, 72°C/2 min 30 sec x30,
1.6 mM $MgCl_2$.

HMTL1161-HMTH1357,
94°C/30 sec, 60°C/30 sec, 72°C/2 min 30 sec x5,
94°C/30 sec, 56°C/30 sec, 72°C/2 min 30 sec x5,
94°C/30 sec, 54°C/30 sec, 72°C/2 min 30 sec x25,
72°C/5 min x1,
2 mM $MgCl_2$.

HMTL1445-HMTH1585,
94°C/30 sec, 45°C/30 sec, 72°C/1 min 30 sec x30,
2.6 mM $MgCl_2$.

HMTL1598-HMTH1642,
94°C/30 sec, 55°C/30 sec, 72°C/45 sec x30,
72°C 4 min x1,
2 mM $MgCl_2$.

Appendix 6 Individual ND4/ND5 clone sequences for mutant and normal patterns of semen sample 94-107. Dots(.) indicate synonymous nucleotides to the Cambridge sequence. Italicised and bolded bases have recognised polymorphisms (Wallace *et al.*, 1995) in which this study also find change. Question mark (?) indicates an un-resolved nucleotide. Numbering refers to the numbering assigned by Anderson *et al.* (1981) to the Cambridge sequence.

```
                    1         1         1         1         1         1         1
                    1         1         1         1         1         1         1
                    6         7         7         7         7         7         7
                    9         0         1         2         3         4         5
                    0         0         0         0         0         0         0
Cambridge           cttcaccggcgcagtcattctcataatcgcccacgggcttacatcctcattactattctgcctagcaaact
178-1-601-3         .........................g..........a.................................
178-1-531-4         .........................g..........a.................................
178-1-531-10        .........................g..........a.................................
178-2-890-13        ...........................................................................
178-2-890-6         .........................g..........a.................................
178-4-531-7         .........................g..........a.................................

                    1         1         1         1         1         1         1
                    1         1         1         1         1         1         1
                    7         7         7         7         8         8         8
                    6         7         8         9         0         1         2
                    0         0         0         0         0         0         0
Cambridge           caaactacgaacgcactcacagtcgcatcataatcctctctcaaggacttcaaactctactcccactaata
178-1-601-3         ............t..c.................t.....................c..............
178-1-531-4         ............t..c.................t.....................c..............
178-1-531-10        ............t...................t.....................c..............
178-2-890-13        ...........................................................................
178-2-890-6         ............t..c.................t.....................c..............
178-4-531-7         ............t..c.................t.....................c..............

                    1         1         1         1         1         1         1
                    1         1         1         1         1         1         1
                    8         8         8         8         8         8         8
                    3         4         5         6         7         8         9
                    0         0         0         0         0         0         0
Cambridge           gctttttgatgacttctagcaagcctcgctaacctcgccttacccccactattaacctactgggagaact
178-1-601-3         ..c.......................a....t.............................a.......c
178-1-531-4         ..c.......................a....t.............................a........
178-1-531-10        ..c.......................a....t.............................a........
178-2-890-13        ...........................................................................
178-2-890-6         ..c.......................a....t.............................a........
178-4-531-7         ..c.......................a....t..........
178-4-388-10        ...........................................................................
178-4-388-11        ...........................................................................

                    1         1         1         1         1         1         1
                    1         1         1         1         1         1         1
                    9         9         9         9         9         9         9
                    0         1         2         3         4         5         6
                    0         0         0         0         0         0         0
Cambridge           ctctgtgctagtaaccacgttctcctgatcaaatatcactctcctacttacaggactcaacatactagtca
178-1-601-3         ..................
178-1-531-4         ..................a.....................................................a...
178-1-531-10        ..................a.....................................................t..a...
178-2-890-13        ...........................................................
178-2-890-6         ..................a..
178-4-388-10        ...........................................................................
178-4-388-11        ...........................................................................

                    1         1         1         1         1         1         1
                    1         1         1         2         2         2         2
                    9         9         9         0         0         0         0
                    7         8         9         0         1         2         3
                    0         0         0         0         0         0         0
Cambridge           cagccctatactccctctacatatttaccacaacacaatgggctcactcacccaccacattaacaacata
178-1-601-3                                                 ...........................
178-1-531-4         .....................................a.....g....g...................
178-1-531-10        .....................................a.....g....g...................
178-2-890-13                                                ...........................
178-4-388-10        ...........................................................................
178-4-388-11        ...........................................................................
```

```
                  1         1         1         1         1         1         1
                  2         2         2         2         2         2         2
                  0         0         0         0         0         0         1
                  4         5         6         7         8         9         0
                  0         0         0         0         0         0         0
Cambridge         aaaccctcattcacacgagaaaacaccctcatgttcatacacctatcccccattctcctcctatcccctcaa
178-1-601-3       ...........................................................................
178-1-531-4       ........................t..........................t..c.....................
178-1-531-10      ........................t..........................t..c.....................
178-2-890-13      ...........................................................................
178-4-531-7                                                          ...t..c.....................
178-4-388-10      ...........................................................................
178-4-388-11      ...........................................................................

                  1         1         1         1         1         1         1
                  2         2         2         2         2         2         2
                  1         1         1         1         1         1         1
                  1         2         3         4         5         6         7
                  0         0         0         0         0         0         0
Cambridge         ccccgacatcattaccgggttttcctcttgtaaatatagtttaaccaaaacatcagattgtgaatctgaca
178-1-601-3       ...........................................................................
178-1-531-4       ......t.....................c.............................t.................
178-1-531-10      ......t.....................c...........................................
178-2-890-13      ............................c...........................................
178-4-531-7       ......t.....................c...........................................
178-4-388-10      ...........................................................................
178-4-388-11      ...........................................................................

                  1         1         1         1         1         1         1         1
                  2         2         2         2         2         2         2         2
                  1         1         2         2         2         2         2         2
                  8         9         0         1         2         3         4         5
                  0         0         0         0         0         0         0         0
Cambridge         acagaggcttacgacccctttatttaccgagaaagctcacaagaactgctaactcatgccccccatgtctaac
178-1-601-3       ...........................................................at.............
178-1-531-4       .........c..................................
178-1-531-10      .........c..................................
178-2-890-13      .........c..................................
178-4-531-7       .........c..................................
178-4-359-9                                    .....t.................at.....c.......
178-4-359-3                                    .....t.................at.............
178-4-388-10      ....................................
178-4-388-11      ....................................

                        1         1         1         1         1         1         1
                        2         2         2         2         2         2         2
                        2         2         2         2         3         3         3
                        6         7         8         9         0         1         2
                        0         0         0         0         0         0         0
Cambridge         aacatggctttctcaacttttaaaggataacagctatccattggtcttaggccccaaaaattttggtgcaa
178-1-601-3       ................................
178-2-890-6                                                                   ..........
178-4-359-9       ................................c....g.............................
178-4-359-3       ................................c..c.g.............................

                        1         1         1         1         1         1         1
                        2         2         2         2         2         2         2
                        3         3         3         3         3         3         3
                        3         4         5         6         7         8         9
                        0         0         0         0         0         0         0
Cambridge         ctccaaataaaagtaataaccatgcacactactataaccaccctaaccctgacttccctaattcccccat
178-2-890-6       ......................t..g........c........g....a......t..........t..
178-4-359-9       ......................t..g........c........g....a......t..........t..
178-4-359-3       ......................t..g........c........g....a......t..........t..

                      1         1         1         1         1         1         1
                      2         2         2         2         2         2         2
                      4         4         4         4         4         4         4
                      0         1         2         3         4         5         6
                      0         0         0         0         0         0         0
Cambridge         ccttaccaccctcgttaaccctaacaaaaaaaaactcataccccccattatgtaaaatccattgtcgcatcca
178-2-890-6       ............a..........t..............t........a............a.........
178-4-359-9       .........              ..t..............t.............................
178-4-359-3       ............a..........t..............t........a............a.........
```

```
                    1         1         1         1         1         1         1
                    2         2         2         2         2         2         2
                    4         4         4         5         5         5         5
                    7         8         9         0         1         2         3
                    0         0         0         0         0         0         0
Cambridge           cctttattatcagtctcttccccacaacaatattcatgtgcctagaccaagaagttattatctcgaactga
178-2-890-6         ..c..g....t.........................a.a..............c........a......
178-4-359-9         ............................................................................
178-4-359-3         ..c..g....t.........................a.a..............c........a......

                    1         1         1         1
                    2         2         2         2
                    5         5         5         5
                    4         5         6         7
                    0         0         0         0
Cambridge           cactgagccacaacccaaacaacccagctctccctaag
178-2-890-6         .....g..a.................a......?....
178-4-359-9         ..............................?....
178-4-359-3         .....g..a.................a......?....
```

Appendix 7 Individual HV1 clone sequences for mutant and normal patterns of semen sample 94-107. Dots(.) indicate synonymous nucleotides to the Cambridge sequence. Numbering refers to the numbering assigned by Anderson *et al.* (1981) and is the numbering of the Cambridge sequence only. Dashes (-) are present for alignment purposes only, but may indicate possible sites of inserts and deletions. Maximum homology alignment was performed using the EMBL Homology Alignment Program in DNASIS.

```
                 1         1         1         1         1         1         1
                 6         6         6         6         6         6         6
                 0         0         0         0         0         0         0
                 3         4         5         6         7         8         9
                 0         0         0         0         0         0         0
Cambridge        gttctttcatggggaagcagatttgggtaccacccaagtattgactcacccatcaacaaccgctatgtatt
180-9-13         ..at.c........g...cat.........a..........c...c.....ca.c.t..-ttt...a...a
180-9-16         ..at.c........g...cat.........a..........c...c.....ca.c.t..-ttt...ag..a
180-9-17         ..at.ck.......g...cat.........a..........c...c.....ca.c.t..-ttt...at..a
180-8-14         ..at.c........g...cat.........a..........c...c.....ca.c.t..-ttt...ag..a
180-8-20         ..at.c........g...cat.........a..........c...c.....ca.c.t..-ttt...a...a
180-9-25         ..at.c.....a..g...cat.........a..........c...c.....ca.c.t..-ttt...ag..a
180-8-30         ............................................................................
180-8-26         ............................................................................
180-9-27         ............................................................................

                 1         1         1         1         1         1         1
                 6         6         6         6         6         6         6
                 1         1         1         1         1         1         1
                 0         1         2         3         4         5         6
                 0         0         0         0         0         0         0
Cambridge        tcgtacattactgccagccaccatgaatattgtacggtaccataaatacttgaccacctgtagtacata-a
180-9-13         ....ga...t....t................a..ta....t....t.g...a..tgtaca.........tc
180-9-16         ....ga...t....t................a..ta....t....t.g...a..tgtaca.........tc
180-9-17         ....ga...t....t................a..ta....t....t.g...a..tgtaca.........tc
180-8-14         ....ga...t....t................a..ta....t....t.g...a..tgtaca.........tc
180-8-20         ....ga...t....t................a..ta....t....t.g...a..tgtaca.........tc
180-9-25         ....ga...t....t................a..ta....t....t.g...a..tgtaca.........tc
180-8-30         ............................................................................-.
180-8-26         ............................................................................-.
180-9-27         ............................................................................-.

                 1         1         1         1         1         1         1
                 6         6         6         6         6         6         6
                 1         1         1         2         2         2         2
                 7         8         9         0         1         2         3
                 0         0         0         0         0         0         0
Cambridge        aaacccaa-tccacatcaaaacccccctcccccatgcttacaagcaagtac-agcaatcaaccctcaactatc
180-9-13         ....ga..c..a...a...ct..t..c.a.g.c.............g..t..a.ta.cttaact...gg.a
180-9-16         ....ga..c..a...a...ct..t..c.a.g.c.............g..t..a.ta.cttaact...gg.a
180-9-17         ....ga..c..a...a...ct..t..c.a.g.c.............g..t..a.ta.cttaact...gg.a
180-8-14         ....ga..c..a...a...ct..t..c.a.g.c.............g..t..a.ta.cttaact...gg.a
180-8-20         ....ga..c..a...a...ct..t..c.a.g.c.............g..t..a.ta.cttaact...gg.a
180-9-25         ....ga..c..a...a...ct..t..c.a.g.c.............g..t..a.ta.cttaact...gg.a
180-8-30         ........-...................................................-..........
180-8-26         ........-...................................................-..........
180-9-27         ........-...................................................-..........

                 1         1         1         1         1         1         1
                 6         6         6         6         6         6         6
                 2         2         2         2         2         2         3
                 4         5         6         7         8         9         0
                 0         0         0         0         0         0         0
Cambridge        acacatcaactgca-actcc-aaagccacccctcacccactaggataccaacaaac-ctacccacccttaa
180-9-13         ......t..acc..cc....a....aa..tt....a....a..cg......a....a...tat.t....--
180-9-16         ......t..acc..cc....a....aa..tt....a....a..cg......a....a...tat.t....--
180-9-17         ......t..acc..cc....a....aa..tt....a....a..cg......a....a...tat.t....--
180-8-14         ......t..acc..cc....a....aa..tt....a....a..cg......a....a...tat.t....--
180-8-20         ......t..acc..cc....a....aa..tt....a....a..cg......a....a...tat.t....--
180-9-25         ......t..acc..cc....a....aa..tt....a....a..cg......a....a...tat.t....--
180-8-30         ............-.....-.....................................-..............
180-8-26         ............-.....-.....................................-..............
180-9-27         ............-.....-.....................................-..............
```

```
                    1         1         1         1         1         1         1
                    6         6         6         6         6         6         6
                    3         3         3         3         3         3         3
                    1         2         3         4         5         6         7
                    0         0         0         0         0         0         0
Cambridge           cagtacatagtacataaagccatttaccgtacatagcacattacagtcaaatcccttctcgtccccatgga
180-9-13            g.............ct.at.gc.c.tg...g.........c....t...caag.c......aat.c...
180-9-16            g.............ct.at.gc.c.tg...g.........c....t...caag.c......aat.c...
180-9-17            g.............ct.at.gc.c.tg...g.........c....t...caag.c......aat.c...
180-8-14            g.............ct.at.gc.c.tg...g.........c....t...caag.c......aat.c...
180-8-20            g.............ct.at.gc.c.tg...g.........c....t...caag.c......aat.c...
180-9-25            g.............ct.at.gc.c.t....g.........c....t...caag.c......aat.c...
180-8-30            ..................................................t.................
180-8-26            ..................................................t.................
180-9-27            ..................................................t.................

                    1         1
                    6         6
                    3         3
                    8         9
                    0         0
Cambridge           tgacccccctcagataggggtcccttga
180-9-13            .-.t.............t.....g...t
180-9-16            .-.t.............t.....g...t
180-9-17            .-.t.............t.....g...t
180-8-14            .-.t.............t.....g...t
180-8-20            .-.t.............t.....g...t
180-9-25            .-.t.............t.....g...t
180-8-30            ............................
180-8-26            ............................
180-9-27            ............................
```

Appendix 8 ND4/ND5 primate homology alignment. *P. troglodytes, G. gorilla, Pongo pygmaeus* and *Hylobytes lar* sequences and alignment follows that of Brown *et al.* (1982). *P. paniscus* sequence is from Genebank (accession no. D38116). Dots(.) indicate synonymous nucleotides to the Cambridge sequence. Italicised and bolded bases have recognised polymorphisms (Wallace *et al.*, 1995) in which this study also find change. Question mark (?) indicates an un-resolved nucleotide. Numbering refers to the numbering assigned by Anderson *et al.* (1981) to the Cambridge sequence.

```
                    1         1         1         1         1         1         1
                    1         1         1         1         1         1         1
                    6         7         7         7         7         7         7
                    9         0         1         2         3         4         5
                    0         0         0         0         0         0         0
Cambridge           cttcaccggcgcagtcattctcataatcgcccacgggcttacatcctcattactattctgcctagcaaact
94-107 mutant       ........................g.........a.....................................
P. troglodytes      .............a.t..c.................a...............t.....................
P. paniscus         ....................t..............a.....................................
G. gorilla          ..............tg....t.....t........a........a......t.....................
P. pygmaeus         ............ac..cc.....g..t.....t..a..c........cc....g.....................
H. lar              ...t..a..t...ac.g.c.................a..a..c..t..cc.g..........t.......

                    1         1         1         1         1         1         1
                    1         1         1         1         1         1         1
                    7         7         7         7         8         8         8
                    6         7         8         9         0         1         2
                    0         0         0         0         0         0         0
Cambridge           caaactacgaacgcactcacagtcgcatcataatcctctctcaaggacttcaaactctactcccactaata
94-107 mutant       .............t..c.................t.....................c...............
P. troglodytes      ....t..t.......c.................t.....c.................................
P. paniscus         ....t..........c.................t.....c.................................
G. gorilla          .............a..c.....c...........t..............c.....c...............
P. pygmaeus         .............a..c.....c.....................c.................c.......
H. lar              .............a........c..............a....g...g..c...g.ct..........g...

                    1         1         1         1         1         1         1
                    1         1         1         1         1         1         1
                    8         8         8         8         8         8         8
                    3         4         5         6         7         8         9
                    0         0         0         0         0         0         0
Cambridge           gctttttgatgacttctagcaagcctcgctaacctcgccttacccccccactattaacctactgggagaact
94-107 mutant       ..c.....................a....t.............................a.......c
P. troglodytes      ..c...........c.....................c.......t..c.....t..c..a..g.....
P. paniscus         ..c...........c.............c.........c.........c.....t..c..a........
G. gorilla          ..cc..............g...........c.....................c...........a.....g..
P. pygmaeus         ..cc.c.....................a.......t...c....a.....c..c.....t..a........
H. lar              ..c..c.........cgc.....................c.................c..a..t.....

                    1         1         1         1         1         1         1
                    1         1         1         1         1         1         1
                    9         9         9         9         9         9         9
                    0         1         2         3         4         5         6
                    0         0         0         0         0         0         0
Cambridge           ctctgtgctagtaaccacgttctcctgatcaaatatcactctcctacttacaggactcaacatactagtca
94-107 mutant       ................a.....................................................a...
P. troglodytes      ...c............t.a................c.............c......t...........a...
P. paniscus         ...c............t.a................ct..c........c......t...........a.t.
G. gorilla          ...c..a...........a..............c.c...c..tt...........tct.........a.t.
P. pygmaeus         ...c..a...a..g....ta.....t.....t..c.....ca.....a.....................a...
H. lar              ..tc..a...a.gg..t.c........gg....c.ct...a.tac...c..c..g......g.....a...

                    1         1         1         1         1         1         1
                    1         1         1         2         2         2         2
                    9         9         9         0         0         0         0
                    7         8         9         0         1         2         3
                    0         0         0         0         0         0         0
Cambridge           cagccctatactccctctacatatttaccacaacacaatgggggctcactcacccaccacattaacaacata
94-107 mutant       .....................................a.....g....g.....................
P. troglodytes      .......g.............g.................a.........................t......
P. paniscus         .......g.............g.................a.........................t......
G. gorilla          .......g........t..t.....................a...c.......a.........c.c.......
P. pygmaeus         ..a.........t.....t.....c.............c.a..ta...c...a........c.........
H. lar              .g.............t.........t..t.......c.a...a....t..a...........a.......
```

```
                    1         1         1         1         1         1         1
                    2         2         2         2         2         2         2
                    0         0         0         0         0         0         1
                    4         5         6         7         8         9         0
                    0         0         0         0         0         0         0
Cambridge           aaaccctcattcacacgagaaaacaccctcatgttcatacacctatcccccattctcctcctatccctcaa
94-107 mutant       ........................t.....................t..c.................
P. troglodytes      ..g...................t..t.....a..tt..............c.....t...........
P. paniscus         ........................t.....a..t.................c.................
G. gorilla          ..........t.............t......a.....g.............c.................
P. pygmaeus         .....t..t........c.....t.........c.................c......t..........
H. lar              .........c..............tat.a..ac.t..g....c..t....c.c.........a.......

                    1         1         1         1         1         1         1
                    2         2         2         2         2         2         2
                    1         1         1         1         1         1         1
                    1         2         3         4         5         6         7
                    0         0         0         0         0         0
Cambridge           ccccgacatcattaccgggttttcctcttgtaaatatagtttaaccaaaacatcagattgtgaatctgaca
94-107 mutant       ......t.....................c.....................................
P. troglodytes      t..t..t.....c..t..a..ca....c.....................................
P. paniscus         t..t..t..t..c..t..a..ca....c.....................................
G. gorilla          ......t..t..c........ca....c.....................................t.
P. pygmaeus         ....ag......cg.t.....cg...ac.....................t.............a.t.
H. lar              ...ta..........t..c...a.tc.c......c.........t........t............a...

                    1         1         1         1         1         1         1         1
                    2         2         2         2         2         2         2         2
                    1         1         2         2         2         2         2         2
                    8         9         0         1         2         3         4         5
                    0         0         0         0         0         0         0         0
Cambridge           acagaggcttacgacccttatttaccgagaaagctcacaagaactgctaactcatgccccccatgtctaac
94-107 mutant       .........c.............................t.................at.............
P. troglodytes      .........c.............................t.t............t....at.......c.....
P. paniscus         .........c.............................t.t............t....at.......c.....
G. gorilla          .........c..a..........................gt....g............a.....g..ct....
P. pygmaeus         .t..g.c.cc..a.....................................tca.t-......g....
H. lar              .t.......cgaa...t...gc.............c.................ctat.......a....

                          1         1         1         1         1         1         1
                          2         2         2         2         2         2         2
                          2         2         2         2         3         3         3
                          6         7         8         9         0         1         2
                          0         0         0         0         0         0         0
Cambridge                 aacatggctttctcaacttttaaaggataacagctatccattggtcttaggccccaaaaattttggtgcaa
94-107 mutant             ..................................c....g.............................
P. troglodytes            ..................................c....g.............................
P. paniscus               ..................................c....g.............................
G. gorilla                ..................................................a.................
P. pygmaeus               ..............g.....................c..........at.................
H. lar                    ..................................................a.................

                          1         1         1         1         1         1         1
                          2         2         2         2         2         2         2
                          3         3         3         3         3         3         3
                          3         4         5         6         7         8         9
                          0         0         0         0         0         0         0
Cambridge                 ctccaaataaaagtaataaccatgcacactactataaccaccctaaccctgacttccctaattcccccccat
94-107 mutant             ........................t..g........c........g....a......t..........t..
P. troglodytes            ........................t.t.....c.........t.......a...c..t......t......
P. paniscus               ........................t.t.....c.........t.......a...c..t......t......
G. gorilla                ....................t...t..g....c.........t..g....a......t..........t..
P. pygmaeus               ................c.g.....ttt..c..c.....tg....c...t.a...........c........
H. lar                    ..................g.a...t....c..c...g...tt.....g..a..c..................

                          1         1         1         1         1         1         1
                          2         2         2         2         2         2         2
                          4         4         4         4         4         4         4
                          0         1         2         3         4         5         6
                          0         0         0         0         0         0         0
Cambridge                 ccttaccaccctcgttaaccctaacaaaaaaaactcatacccccattatgtaaaatccattgtcgcatcca
94-107 mutant             ............a.........t.............t........a............a.........
P. troglodytes            ...c.........a.........................t...........g.........a....g....
P. paniscus               ...c.........a.........................t...........g.........a....g....
G. gorilla                ..........t..a.c..t.............g..............c........t..c..........
P. pygmaeus               taccg.t......a.......c.............c.........c.........a.ggcca........g
H. lar                    tacag.......ta.......c..t.....g....t......g..c..c......ttg.cca.t..c..t.
```

	1	1	1	1	1	1	1
	2	2	2	2	2	2	2
	4	4	4	5	5	5	5
	7	8	9	0	1	2	3
	0	0	0	0	0	0	0
Cambridge	cctttattatcagtctcttccccacaacaatattcatgtgcctagaccaagaagttattatctcgaactga						
94-107 mutant	..c..g....t.....a.a.....c.....a.....						
P. troglodytes	.....c..t..c..t.....a.....c.....a.....g						
P. paniscus	.....c..t..c..t.....a.....c.....a.....g						
G. gorilla	.....c.....c.....tc.a.....c.....a.g....						
P. pygmaeus	.....c.....c..ta....a.....t..c.....ga.....acc..cg..a.a.....						
H. lar	.....a.....c..a..t.....t.....ac.....acc.....t..a.....						

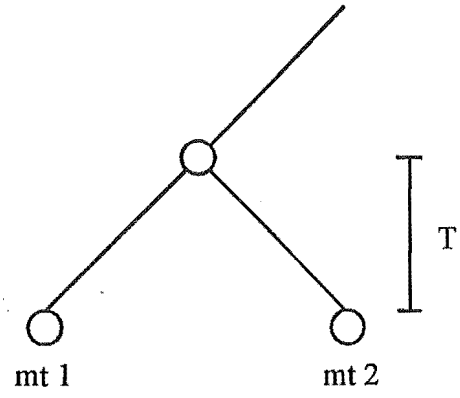
	1	1	1	1
	2	2	2	2
	5	5	5	5
	4	5	6	7
	0	0	0	0
Cambridge	cactgagccacaacccaaacaacccagctctccctaag			
94-107 mutant	.....g..a.....a.....?....			
P. troglodytes	.....a.....			
P. paniscus	.....a.....			
G. gorilla	.....a.....tt..a.....			
P. pygmaeus	tg....a.a..c.....g...cta..a.....a.....			
H. lar	.....a.tg.....gctag.a.....			

Appendix 9. An evaluation of the derivation of k_{nuc} from Kimura, (1981) as it applies to mitochondrial pseudogenes.

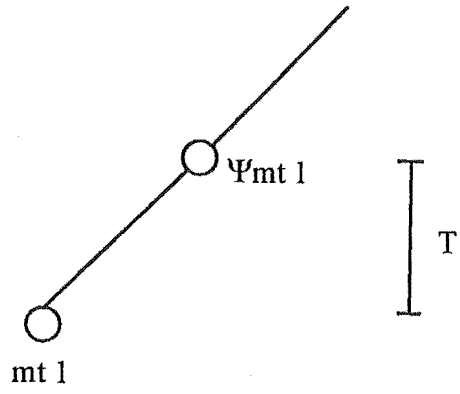
Kimura (1981), states that the base substitution rate per unit time(k_{nuc}) is equal to the overall corrected base substitution rate (K) divided by twice the divergence time (T) i.e. :

$$k_{nuc} = K/2T \quad (1)$$

This is based on dendrogram relationships of terminal taxa where the nucleotide changes in time T are subdivided into two internode branches from an internal node to the terminal taxa. i.e. :



However, for mtDNA pseudogenes, the rate of change along the nuclear branch since the time of insertion approximates zero. The terminal pseudogene node therefore approximates the internal node (assuming branches represent nucleotide changes) and therefore the tree can be re-drawn. i.e. :



Thus to better approximate the base substitution rate per unit time when applied to mtDNA pseudogenes equation (1) should be modified to :

$$k_{nuc} = K/T \quad (2)$$

as changes are only occuring along one internode branch in time T.